(54) Title: SMALL HAIRPIN RNA LIBRARIES

(57) Abstract: The present invention relates to novel and rapid methods of identifying genes and pathways in mammalian cells based on RNA interference (RNAi). The invention includes methods of producing and using libraries of small interfering RNAs known as short hairpin RNAs (shRNAs). The dsRNAs generated by the methods of the invention can be introduced into cells in a number of ways, including as an isolated nucleic acid, (e.g., chemically synthesized or translated in vitro), or by transfection of an expression vector (e.g., a plasmid) that includes sequence encoding the shRNA.

SMALL HAIRPIN RNA LIBRARIES

This application claims the benefit of U.S. Serial No. 60/500,860, filed September 5, 2003. For the purpose of any U.S. patent(s) that may issue from the present application, the contents of the priority document are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

This invention relates generally to the production and use of collections of nucleic acid molecules, particularly collections of small hairpin RNAs (shRNAs) and shRNA expression templates.

BACKGROUND

5   Experimental models have been developed for many human disorders, including neurodegenerative diseases such as Huntington's disease (HD) and amyotrophic lateral sclerosis (ALS). For example, cells that have been genetically altered to express disease-causing mutant genes (*e.g.*, huntingtin, SOD1, or alpha-synuclein) have been studied in cell

10   culture and in animal models. These models have been used to determine the molecular mechanisms of disease pathophysiology, and they provide tractable systems to test a range of hypotheses. Most phenotypes assayed in such models develop rapidly, and the cells provide an easily accessible source of proteins, lipids, and nucleic acids. Cell-based high-throughput screens make it possible to use such models to identify potential therapeutic compounds.

15   A compelling approach that is possible in some small organisms is to screen for suppressor or enhancer mutations that alter the disease phenotype, thus defining genes and pathways that are critical in the development and progression of the disease. Traditionally, this type of genetic analysis has not been widely used in mammals and mammalian cell models because of technical difficulties associated with mutagenesis in mammalian systems.

20   Fortunately, recently refined RNA interference (RNAi) techniques now permit feasible loss-of-function analysis in mammalian cells.

Small interfering RNA (siRNA) molecules can selectively suppress gene expression through RNAi (*see* Hannon, *Nature* 418:244-251, 2002). The suppression is predominantly a cytoplasmic, post-transcriptional event that is evolutionarily conserved (*see* Hutvagner and

25   Zamore, *Curr. Opin. Genet. Dev.* 12:225-232, 2002). While the phenomenon was well documented in lower eukaryotic organisms, the long dsRNA used for RNAi in those systems evoked a potent interferon response in mammalian cells. This response is not triggered when

1

siRNAs shorter than 30 base pairs are used (Elbashir *et al.*, *Nature* 411:494-498, 2002).

The silencing mechanism is not fully understood. Current models suggest that dsRNA is processed into siRNA by the RNAse III enzyme, Dicer. The siRNA then participates in a complex called RNA-induced silencing complex (RISC). This complex then interacts with the target mRNA, which is then cleaved and degraded (Martinez *et al.*, *Cell* 110:563-574, 2002; Schwarz *et al.*, *Mol. Cell.* 10:537-548, 2002).

Initial experiments in mammalian cells used chemically synthesized RNA oligonucleotides (*see* Elbashir *et al.*, *supra*). Subsequently, researchers demonstrated that plasmids could encode short hairpin-forming RNAs (shRNAs) that could function as siRNAs (*see* Sui *et al.*, *Proc. Natl. Acad. Sci. USA* 99:5515-5520, 2002; Brummelkamp *et al.*, *Science* 296:550-553, 2002). These plasmids consist of an RNA polymerase III promoter, followed by sequence encoding a "sense" sequence, a loop structure, and an "antisense" sequence. A variety of loops have been successfully employed, as have several polymerase III promoters (Brummelkamp *et al.*, *Science* 296:550-553, 2002; Miyagishi and Taira, *Nucl. Acids Res. Suppl.* 2:113-114, 2002). U6-promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells (Paddison *et al.*, *Nature Biotechnol.* 20:497-500, 2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells (Paul *et al.*, *Genes & Dev.* 16:948-958, 2002), and siRNAs can be effectively expressed in human cells (Sui *et al.*, *Nature Biotechnol.* 20:505-508, 2002). A DNA vector-based RNAi technology to suppress gene expression in mammalian cells is described by Yu and Turner (*Proc. Natl. Acad. Sci. USA* 99:5515-5520, 2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells has also been described (*Proc. Natl. Acad. Sci. USA* 99:6047-6052, 2002). The ability to encode siRNA in expression vectors such as plasmids permits stable repression of specific genes.

SUMMARY

The present invention features, *inter alia*, methods for producing collections of shRNAs, which we may refer to as shRNA libraries. The shRNAs can be produced using shRNA expression templates made from mRNAs of known sequence or from mRNAs within one or more pools, populations, or sets whose sequences are incompletely understood (*e.g.*, a pool of mRNA in which the sequence of one or more of the mRNAs is less than fully known). The invention also features the libraries *per se,* methods for using those libraries, and the primers, agents (*e.g.*, tagged nucleic acids (*e.g.*, biotinylated nucleic acids)), vectors,

2

cells, and unique intermediate constructs prepared in the course of generating the shRNA libraries. For example, the invention features a composition that includes a plurality of cDNA tags (*e.g.*, at least 100 cDNA tags). The sequence of at least one of the cDNA tags can be at least partially unknown. Kits, including any combination of the various compositions
5   described herein, are also within the scope of the invention.

The methods of shRNA library construction can include providing a cDNA library (*e.g.*, a normalized cDNA library), generating at least one shRNA expression template from each of a plurality of cDNAs in that library, and transcribing shRNAs from the shRNA expression template(s). Preferably, shRNA expression templates are made from a majority of
10   the cDNAs in the cDNA library (*e.g.*, from at least (or about) 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, or 99% or more of the cDNAs in the library). Thus, shRNA expression templates can be made from all or essentially all of the cDNAs. The complexity of the cDNA library can vary, and the collection of shRNA expression templates and, ultimately, the shRNAs produced may vary accordingly. While the shRNA library can be small (*e.g.*, having about
15   100 members), complex shRNA libraries can include up to about 1-2 x 10^4 shRNAs, and in all relevant aspects of the invention (*e.g.*, in using shRNA libraries to identify genes that are, or that encode, therapeutic targets or candidate therapeutic targets), one can employ more than one shRNA library (*e.g.*, screening methods can employ 2-5 or more shRNA libraries). As noted, shRNAs can be generated without any knowledge (or with incomplete knowledge)
20   of the sequences of the mRNAs from which the shRNA expression templates are ultimately made; we note too that the methods can be carried out without chemical synthesis.

The methods for producing collections of shRNAs (*e.g.*, an shRNA library) can include the steps of providing a plurality of nucleic acid molecules that include shRNA expression templates (*e.g.*, shRNA vector constructs) and transcribing shRNA from the
25   shRNA expression templates, thereby producing a library of shRNAs. The nucleic acids that include the shRNA expression templates and shRNA expression templates *per se* can be obtained in a number of ways, as described herein. For example, one can obtain a set of cDNA tags (as illustrated in, for example, FIGs. 2A and 2C, each cDNA tag has a first end and a second end). Dual hairpin structures can then be generated by ligating a first hairpin
30   loop to the first end of each cDNA tag and a second hairpin loop to the second end of each cDNA tag. Typically, one of the hairpin loops will have a blunt end and the other will have overhanging nucleotides (this facilitates directionality). In addition, one of the hairpin loops can include a cleavage site, and one can use that site to facilitate opening the dual hairpin structures (exposure to heat can also facilitate opening by denaturing the bonds between the

3

nucleotides). When cleaved and denatured, the hairpin structures may be referred to as linearized, single-stranded dual hairpin structures. The shRNA expression templates are then produced by synthesizing, on the linearized, single-stranded dual hairpin structures, a second, complementary strand of DNA (*see* FIGs. 2A-2D).

5   As shown in FIGs. 3A and 3B, shRNA expression templates can also be made by providing a plurality of vectors each of which includes a promoter that is operably linked to an insert that includes the sense strand of a cDNA tag and a sequence that, when transcribed, will produce a hairpin loop. The insert is transcribed (*e.g.*, in a cell type described herein) and the sequence is extended by a self-priming reaction. This reaction generates a sequence

10   that is complimentary to the sense strand of the cDNA tags, thereby producing a stem-loop-stem structure (see the lower half of FIG. 3A). Denaturing the bonds between the nucleic acids along the stem of the stem-loop-stem structure (with, for example, heat or a chemical agent) produces a denatured construct, and one can synthesize, one that single stranded construct, a second strand that is complimentary to the sequence of the denatured construct.

15   This process generates shRNA expression templates (as shown, with dual T7 promoters, in FIG. 3B).

         Where it is necessary to transcribe shRNA from the shRNA expression templates, one can modify the shRNA expression templates by (a) operably linking the shRNA expression templates to a promoter and (b), optionally, removing a portion of the sequence of the first

20   hairpin loop (*e.g.*, the sequence between two restriction sites, such as the *Bcg I* sites shown in FIG. 3C). The cell types described herein can then be transfected with the shRNA expression construct. In another embodiment, shRNAs can be transcribed from the shRNA expression templates by inserting the shRNA expression templates into vectors (*e.g.*, plasmid vectors), thereby producing an shRNA vector construct, and transfecting cells with the shRNA vector

25   construct. Optionally, one can remove a portion of the sequence of the first hairpin loop.

         Methods for producing a set of cDNA tags are also within the scope of the invention, and those methods (like the methods to produce shRNA expression templates) can be carried out in the course of generating an shRNA library. cDNA tags can be obtained by providing a cDNA library and exposing members of the cDNA library to at least two restriction enzymes.

30   The enzymes are selected to cleave the members of the cDNA library into fragments 10-50 nucleotides long (*i.e.*, fragments having a sense strand that is 10-50 nucleotides long and an antisense strand that is 10-50 nucleotides long). For example, the fragments that constitute the cDNA tags can be 14, 16, 18, 20, 22, 24, 26, 28, or 30 nucleotides long.

4

In a specific embodiment, illustrated in FIGs. 1A and 1B or 1C, a set of cDNA tags can be obtained by: (a) providing a cDNA library, the members of which have been modified to include, at a first terminus, an overhanging sequence representing a cleaved restriction site; (b) ligating members of the library to a first linker that includes (i) an overhanging sequence complimentary to the overhanging sequence at the first terminus, thereby reconstituting a first restriction site, and (ii) a first immobilization agent, thereby generating ligated members; (c) immobilizing the ligated members by exposing the first immobilization agent to a substrate-bound partner, thereby generating ligated, substrate-bound members; (d) exposing the ligated, substrate-bound members to a first restriction enzyme that cleaves the ligated, substrate-bound members at a second restriction site, thereby generating a restriction-cut second terminus on the ligated, substrate-bound members; (e) exposing the ligated, substrate-bound members to a second restriction enzyme that cleaves the first restriction site, thereby generating freed members; (f) ligating the freed members to a second linker comprising (i) an overhanging sequence complimentary to the restriction-cut second terminus, (ii) a type IIS restriction site and (iii) a second immobilization agent, thereby generating second linker-bound members; (g) exposing the second linker-bound members to a restriction enzyme that recognizes the type IIS restriction site and, by cleaving the second linker-bound members, generates a new first terminus; (h) immobilizing the second linker-bound members by exposing the second immobilization agent to a substrate-bound partner, thereby generating final substrate-bound members; and (i) exposing the final substrate-bound members to the first restriction enzyme, thereby generating the set of cDNA tags.

In any of the methods described herein, the mRNA (or a cDNA library produced from that mRNA) can, but does not necessarily, contain one or more members having sequences that are incompletely known, and the cDNA library can be normalized or subtracted (or both normalized and subtracted). Moreover, the mRNA (or a cDNA library produced from that mRNA) can be obtained from essentially any biological source of mRNA. Suitable sources include all animal (*e.g.*, mammalian) tissues and cell types, including human tissue, whether obtained from a subject who is considered healthy or from a diseased (*e.g.*, cancerous or inflamed) tissue. Other suitable sources include non-human primates (*e.g.*, monkeys, apes, gorillas, and chimpanzees), animals commonly used in medical research (*e.g.*, rodents such as mice, rats, hamsters, and guinea pigs), farm animals (*e.g.*, horses, cows, pigs, sheep, and goats), marine life (*e.g.*, fish, shellfish, dolphins, whales, and the like), animals commonly kept as pets (*e.g.*, dogs, cats, birds, turtles, frogs, and lizards), and invertebrates such as flies (*e.g.*, *Drosophila*) and worms (*e.g.*, *C. elegans*). Plants, bacteria, fungi, and yeast are also

5

suitable sources for the mRNA used in the methods of the present invention. The shRNA
libraries can also be constructed from cell lines, including cancer cell lines (*e.g.*, the HeLa
cell line) and immortalized cells. shRNA libraries obtained using these sources are within the
scope of the present invention. The cell types mentioned here, whether maintained in culture
or *in vivo*, can also be used in the screening methods of the invention to elucidate
biochemical pathways and identify genes that can be targeted by therapeutic agents (*e.g.*, by
siRNAs or other silencing agents; by antibodies, small molecules, and other therapeutic
compounds directed against the polypeptides encoded by the identified genes). More
specifically, the tissue used to make or screen the shRNA libraries can consists of a single
organ or cell type (*e.g.*, a hepatocyte, a fibroblast, a neuron, a glial cell, an epithelial cell, a
myocyte, an adipocyte, a blood cell (*e.g.*, an erythrocyte or leukocyte), an osteoblast,
osteoclast, or other bone-associated cell, or endocrine cell). While differentiated cells are
useful, the methods of the invention can also employ undifferentiated cells (*e.g.*, stem cells
from an embryo or other prenatal animal, an umbilical cord, or an adult) and partially
differentiated cells can also be used. For example, the cells may be obtained at a particular
point in the development of the organism.

In the methods described herein, we refer to a cleavage site within one of the loop
structures used to generate dual hairpin structures and, ultimately, shRNAs. That cleavage
site can include any cleavable entity. For example, the cleavage site can include a pair of
uracil ribonucleic acids (cleavable by uracil glycosylase or a biologically active variant or
fragment thereof) or a restriction enzyme recognition site (cleavable by a restriction enzyme).

Vectors suitable for use in the methods of the invention include plasmids, which may
include one or more regulatory sequences (*e.g.*, promoters and/or enhancers (*e.g.*, an
inducible or constitutively active promoter)). More specifically, the promoter can be (or the
regulatory sequence can include) an RNA polymerase III, RNA polymerase II, U6, T7, SV40,
or H1 promoter. Where two regulatory sequences are used, one can be oriented on each side
of the insert.

The methods of the invention rely on molecular biology techniques including cutting
and ligating nucleic acid sequences. Those of ordinary skill in the art routinely use such
techniques and are capable of adjusting the methods described herein to accommodate
promoters, vectors, hairpin loops, primers, linkers, or other entities that differ from the
specific examples provided herein. For example, in the methods described above, we refer to
a first restriction site, which can be a *BamH* I restriction site, but other restriction sites that
produce overhanging nucleic acids can be readily used. Similarly, the first immobilization

6

agent or the second immobilization agent can be biotin or a polypeptide, but other anchors are known in the art and can be incorporated. Where the first immobilization agent is biotin, the substrate-bound partner can be avidin or streptavidin; where the first immobilization agent is a polypeptide, the substrate-bound partner can be an antibody that specifically binds the polypeptide. The second restriction site referred to in the methods described above can have a four base-pair recognition sequence (*e.g.*, it can be recognized by *Aci* I, *Alu* I, *Bfa* I, *BfuC* I, *BstU* I, *CviJ* I, *CviR* I, *Dpn* I, *Dpn* II, *Fat* I, *Hae* III, *Hha* I, *HinP1* I, *Hpa* II, *Mbo* I, *Mnl* I, *Mse* I, *Msp* I, *Nla* III, *Pho* I, *Rsa* I, *Sau3A* I, *Tai* I, *Taq*ª I, *Tha* I, or *Tsp509* I). Type IIS restriction sites are also employed, and can be an *EcoP14* I, *Eco57* I, *Bpm* I, *BspH614* I, *Bco35* I, *Gsu* I, *Bce83* I, *Bsg* I, or *Mme* I restriction site, and the restriction enzyme that recognizes the type IIS restriction site is of *EcoP14* I, *Eco57* I, *Bpm* I, *BspH614* I, *Bco35* I, *Gsu* I, *Bce83* I, *Bsg* I, or *Mme* I, respectively. While numerous cell types can be used, we note that the step of transcribing shRNA from the shRNA expression templates can be carried out not only in cell culture and *in vivo*, but also in cell-free extracts.

A library made by a method of the invention can be used in any way other shRNA libraries can be used. For example, the library can be used to identify potential therapeutic target genes involved in pathogenesis or disease progression. Accordingly, the invention also includes methods for identifying a target gene using an shRNA library made by a method described herein. A library of the invention can also be used to identify dsRNAs with potential therapeutic action, and those methods are also within the scope of the invention. Libraries (*e.g.*, shRNA libraries or libraries of shRNA vector constructs) made by a process described herein are also within the scope of the invention (*e.g.*, libraries made from cells or organisms for which we have an incomplete understanding of gene expression and/or sequence), and these libraries can be packaged with, optionally, instructions for their use and/or reagents useful in any of the library-based screening methods of the invention. Any of the intermediate constructs described herein can be similarly packaged, and kits containing these constructs are within the scope of the present invention.

In specific embodiments, use of the present libraries includes methods of identifying a therapeutic target for the treatment of a disease by: providing an shRNA library; providing a cell that serves as a model of the disease; contacting the cell with one or more shRNAs from the library; and evaluating an effect of the shRNA on a preselected parameter in the cell. An improvement in the preselected parameter (to, *e.g.*, a statistically significant degree) indicates that the shRNA identifies a therapeutic target. The cell types can be any of those provided herein, include the cell of a simple organism (*e.g.*, *Drosophila*). A single shRNA can be

7

expressed, as can a pool of shRNAs. With respect to the preselected parameter, it can be a

metabolic parameter (*e.g.*, enzyme activity, ionic gradients or concentrations, ligand-receptor

binding, or ligand-receptor activation), a pathophysiological parameter (*e.g.*, apoptosis,

necrosis, proliferation (*e.g.*, uncontrolled or undesirable proliferation), or senescence), a

5   developmental parameter, or a phenotypic parameter (*e.g.*, morphology, motility, or

developmental progression (or lack thereof)). Where enzyme activity is assayed, the enzyme

can be a kinase, protease, helicase, or polymerase, and can be ATP-dependent or ATP-

independent. A change in ion concentration can be measured as a change in ion flux, ion

gradient, plasma membrane potential, mitochondrial potential, or a change in the

10   concentration of a specific ion (*e.g.*, sodium, potassium, calcium, or chloride concentrations).

Any of the screening methods described herein can further include the step of identifying the

shRNA that evokes the cellular response and the gene and gene product it inhibits.

      Libraries of oligonucleotides have been constructed before, but the prior methods

require a substantially complete understanding of gene expression in the cells from which the

15   library is made (*i.e.*, they are possible where gene expression and gene sequence are known).

Prior methods relied on the synthesis of small inhibitory RNAs (siRNAs) directed to the

sequences of each of the genes known to be expressed in a given cell type or organism (*e.g.*,

C. elegans). As the annotation of the human genome improves, one could theoretically

design an individual siRNA for each human mRNA, but this would be a massive undertaking.

20   To generate siRNA activity, one would have to synthesize two primers, both of which contain

at least 60 nucleotides, for each gene. Using a conservative estimate of 30,000 genes in the

genome of most mammalian species, a vast amount of synthesis would be required to

generate a complete RNAi library for a single species. This would clearly consume a

tremendous amount of time and money; so much as to be impractical in many (and probably

25   most) academic research institutions. There are additional limitations to the prior approach.

For example, the libraries would be species-specific and reliant on our knowledge of mRNA

sequences; if a given mRNA were not in a database, no shRNA or siRNA would be generated

against it. The methods and libraries of the present invention obviate these difficulties, as no

prior sequence information or expression data is required. Accordingly, the libraries of the

30   invention include those generated from cells for which expression data is incomplete; the

libraries of the invention can be made without fully understanding which genes are expressed

and/or without having the sequences of expressed genes.

      More specifically, the present methods represent an improvement in library

construction by creating an shRNA library using cDNA. The present methods can be carried

8

out by providing a cDNA library; a normalized library can be used because in most non-normalized cDNA libraries a few transcript species comprise the majority of the population (without normalization, relatively minor transcripts that may be important may be under-represented). dsDNA-hairpin templates are generated from the cDNA, and those templates

5   can be used to directly transcribe a library of shRNAs. Alternatively, the templates can be inserted into a vector for amplification and/or expression (as noted elsewhere, vectors containing such templates are within the scope of the present invention). The shRNAs can be used in a screening assay (*e.g.*, a high-throughput screen), which can be configured to identify genes that are relevant to disease pathophysiology (relevant genes and/or the protein

10   they encode are potential therapeutic targets; accordingly, the methods of the invention may be described as methods for identifying potential therapeutic targets).

An effect on one or more of the cells or simple organisms can indicate that the shRNA is a positive result. In some embodiments, the effect is selected from the group consisting of a change in phenotype (*e.g.*, a change in phenotype selected from the group consisting of a

15   change in morphology, proliferation, movement, development, viability and death); enzyme activity (*e.g.*, a change in enzyme activity selected from the group consisting of a change in kinase, protease, helicase, and polymerase activity, a change in ATP-dependent enzyme activity or independent enzyme activity); ion concentrations (*e.g.*, a change in ion concentrations selected from the group consisting of a change in ion flux, ion gradient,

20   plasma membrane potential, mitochondrial potential, and calcium concentrations); ligand-receptor binding; and ligand-receptor activation. In some embodiments, the effect is a change in a parameter selected from the group consisting of a metabolic parameter, a pathophysiological parameter or a developmental parameter.

The present invention may have one or more advantages. For example, the present

25   methods can be carried out without prior knowledge of the sequence of the genes to be screened. Abundant genes as well as rare and differentially expressed genes (which are likely to be of interest) can be screened using the present methods because of the use of normalized and/or subtracted cDNA libraries as a source of genetic material. Finally, as the methods are compatible with the use of high-throughput screening methods, one can rapidly and

30   accurately screen therapeutic targets and compositions.

Other features and advantages of the invention will be apparent from the drawings, the detailed description, and the claims.

9

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A-1C are schematic diagrams illustrating methods described herein. **FIG. 1A** illustrates the production of a normalized cDNA population from mRNA (shown in further detail in FIG. 4). **FIG. 1B** illustrates methods for producing cDNA tags from the 5' end of the sense strand, and **FIG. 1C** illustrates methods for producing cDNA tags from the 3' end of the sense strand.

FIGs. 2A-2F are diagrams illustrating methods described herein. **FIG. 2A** illustrates methods for producing a sense-loop-antisense shRNA expression construct from a cDNA tag from the 5' end of the sense strand. **FIG. 2B** illustrates methods for producing a sense-loop-antisense shRNA expression construct from a cDNA tag from the 3' end of the sense strand. **FIG. 2C** illustrates methods for producing an antisense-loop-sense shRNA expression construct from a cDNA tag from the 5' end of the sense strand. **FIG. 2D** illustrates methods for producing an antisense-loop-sense shRNA expression construct from a cDNA tag from the 3' end of the sense strand. **FIG. 2E** illustrates methods for producing a sense-loop-antisense shRNA from an shRNA expression construct. **FIG. 2F** illustrates methods for producing an antisense-loop-sense shRNA from an shRNA expression construct.

FIGs. 3A-3C are diagrams illustrating methods described herein. **FIG. 3A** illustrates an alternative method for producing an shRNA expression construct. **FIG. 3B** illustrates an shRNA expression construct made by this alternative method. **FIG. 3C** illustrates an alternative method of inserting an shRNA expression construct into a vector.

FIG. 4 is a schematic diagram illustrating the process of generating a normalized cDNA population (or library), as shown in **FIG. 1A**, in more detail.

DETAILED DESCRIPTION

The present invention relates, in part, to methods of identifying genes and pathways in cells (*e.g.*, mammalian cells, including human cells and cell lines) using methods based on RNA interference (RNAi). The methods described herein can include producing libraries of small interfering RNAs known as short hairpin RNAs (shRNAs). The methods for producing these libraries, and the intermediate constructs formed in the process, are themselves unique and are another aspect of the invention.

The methods of producing shRNA libraries described herein include generating cDNA tags from cDNA libraries, and converting those tags into shRNA expression templates from which shRNA can be produced.

10

cDNA Libraries

mRNA (usually, poly-A$^+$ mRNA) is used to generate cDNA libraries; cDNA libraries suitable for use in the present methods, and methods for making such libraries, are known in the art. One can, therefore, generate shRNAs using a cDNA library that is commercially available or one that is made from mRNA obtained from any animal, tissue, or cell type. If desired, the cDNA library can be normalized; this minimizes the risk that minor transcripts will be under-represented. Normalized libraries are also commercially available (*e.g.*, from ResGen/Invitrogen (Carlsbad, CA) and from Stratagene (La Jolla, CA), *inter alia*), and they can also be made (*e.g.*, normalized libraries can be constructed from a selected source of mRNA by methods known in the art (*see, e.g.*, the protocol described in Carninci *et al.* (*Genome Res.* <u>10</u>:1617-1630, 2000, which is hereby incorporated by reference)). Briefly, to make a normalized cDNA library using this protocol, full-length mRNA is captured with a 5' cap-trapper approach (*Id.*). After first strand cDNA synthesis (*e.g.*, using an oligo-dT primer), the sense RNA strand is removed from the RNA-DNA hybrid by treatment with RNase I. The 3' end of the single, anti-sense-strand DNA is then extended with a dG tail. The resulting population of cDNAs is normalized by mixing with biotinylated mRNA from the original cell population, and abundant mRNA-cDNA species are then eliminated by precipitation using the biotin tag (*e.g.*, with avidin or streptavidin beads). This approach has been shown to significantly reduce the number of abundant transcripts while preserving rare species (*Id.*). Lastly, a dC primer is used to synthesize the second, sense strand of the cDNA. One exemplary method is illustrated in Fig. 4.

Where one desires to make such a cDNA library, one can begin by (a) capturing mRNA (*e.g.*, full-length mRNA) with a 5' cap-trapper approach (to produce 5'-capped mRNA molecules); (b) synthesizing a first strand of cDNA (with, *e.g.*, an oligo-dT primer); (c) removing the sense RNA strand from the RNA-DNA hybrid (with, *e.g.*, RNase I); (d) extending the 3' end of the antisense DNA to include a dG tail (to produce a population of cDNAs); (e) mixing the cDNA with biotinylated mRNA from cells of the same type as were used to obtain the mRNA (to produce biotin-tagged cDNAs); and (f) precipitating the biotin-tagged cDNAs to eliminate abundant mRNA-cDNA species. Once the normalized population is obtained, one can synthesize the second, sense strand of the cDNA (with, *e.g.*, a dC primer).

The cDNA library (whether normalized or not) can be derived from any mRNA source of interest. For example, one can select a cell-specific, tissue-specific, organ-specific, species-specific, or developmental stage- or age-specific source of mRNA; the source can be

11

primary (e.g., from an animal, organ, tissue, or primary cell) or tissue culture (e.g., a cultured cell or tissue). The source can be human or non-human (e.g., a non-human mammal such as a mouse, rat, hamster, guinea pig, cat, dog, horse, cow, goat, pig, sheep or monkey). The cDNA can also be made from an animal or cell that is a disease model (e.g., an animal, tissue,

5   or cell, e.g., a primary cell derived from an animal model or a cell from a tissue culture model). As one example, the disease model can be a model of human neurodegenerative disease.

The libraries can be prepared singly, or in pairs or groups, e.g., using pairs or groups of sources that are the same except for one (or more, but typically only one) major

10   characteristic, such as disease state or stage. One of skill in the art would readily be able to select an appropriate pair or group of sources. For example, cDNA can be prepared from diseased and normal cells (or tissues or animals); pre-cancerous and cancerous cells or tissues; cancerous and metastatic cells; and the like. Pairs or groups of cDNA libraries can also be made from cells, tissues, or animals of different ages or developmental stages. For

15   example, libraries can be made from adolescent, fully adult, and/or aged cells; from any embryonic stage (e.g., E1, E2, E3, E4, E5, E6, etc.); from cells that have (and, optionally, comparable cells that have not) been exposed to a stimulus (e.g., a drug, toxin, metabolite, or vitamin) or an environmental factor (e.g., a stress (e.g., a heat shock) or radiation)); from cells in different states (e.g., different stages of the cell cycle) or at various points of

20   differentiation (e.g., stem cells versus partially differentiated versus terminally differentiated cells); or from other variable states (e.g., quiescent vs. activated; stimulated vs. unstimulated; fed vs. starved; and viable vs. apoptotic). One of ordinary skill in the art would appreciate that there are myriad combinations and states of biological interest that could be examined using the methods described herein. The cells and cell types described here as useful in

25   making the cDNA libraries on which the present methods are based can also be used in the methods of identifying target genes or potential therapeutics.

The source of the cDNA can be a subtracted library (e.g., a subtracted, normalized library). Subtracted libraries are commercially available, and numerous cDNA subtraction methods have been reported. In general, those methods involve hybridization of cDNA from

30   one population (the "tester") to an excess of mRNA (cDNA) from another population (the "driver") followed by separation of the un-hybridized fraction (the "target") from hybridized, common sequences. The latter step is usually accomplished by hydroxyapatite chromatography, avidin-biotin binding, or oligo(dT)30-latex beads. PCR-based cDNA subtraction methods are also known in the art, and include the methods described in

12

Diatchenko *et al.* (*Methods Enzymol.* <u>303</u>:349-80, 1999); Zhumabayeva *et al.* (*Biotechniques* <u>30</u>:512-516, 518-520, 2001); Pacchioni *et al.* (*Biotechniques*. <u>21</u>:644-646, 1996); and Diatchenko *et al.* (*Proc. Natl. Acad. Sci. USA*, <u>93</u>:6025-6030, 1996). Moreover, more than two cDNA libraries can be used (*e.g.*, three or four or more, as described, for example, in WO
5   03/033673). The publications cited in this paragraph are hereby incorporated by reference, and the steps described above can be carried out in the course of making an shRNA library.

As noted above (and as shown, for example, in FIGs. 1A and 4), a normalized cDNA population can be prepared using a "dC primer" that is used to synthesize the second, sense strand of the cDNA. That dC primer can incorporate a sequence recognized by (*i.e.*, cleaved
10   by) a restriction enzyme (any of which may be more specifically referred to as an endonuclease). Thus, in some embodiments, the primer (*e.g.*, the dC primer) used in synthesizing the "sense" strand of cDNA includes a restriction enzyme recognition site (*e.g.*, *BamH* I; *BamH* I is referred to herein simply to illustrate the methods by which the shRNA library can be made). In that case, the resulting double stranded cDNA has the same
15   restriction enzyme site (*e.g.*, *BamH* I) at a position corresponding to the 5' side or 5' end of the original sense strand of mRNA (this intermediate nucleic acid (a double stranded cDNA that includes an extended polynucleotide sequence (*e.g.* a poly-C, poly-G extension) and, further, a restriction site) is also within the scope of the present invention). Many restriction enzymes are known in the art and can be used in creating shRNA libraries (suitable enzymes
20   are described in standard textbooks and can be found on the internet (*e.g.*, on the website of New England Biolabs Inc., Beverly, Mass. USA)).

Once the restriction site is incorporated, the cDNA can be digested with an enzyme that recognizes that site, and a biotinylated anchoring linker can be ligated to the enzyme-modified (*e.g.*, 5') end of the cDNA. The linker will include an anchoring moiety such as
25   biotin (or any other moiety that is part of a binding pair that can be used to anchor or select the attached components) and, linked to the biotin, complementary, paired nucleic acid sequences with one sequence overhanging another. The overhanging "free" sequence in the linker will be complementary to that left exposed by enzymatic digestion. For example, if the cDNA is modified to include a terminal *BamH* I site, and is digested with *BamH* I, the
30   biotinylated anchoring linker used would include overhanging nucleotides, at least some of which are complementary to the overhang left on the cDNA molecule by *BamH* I digestion (see Fig. 1B). Although the present discussion refers to the use of *BamH* I, any restriction enzyme (particularly those with recognition sequences of comparable length, that would cut with approximately the same frequency) can be used in place of *BamH* I.

13

Alternatively, or in addition, another primer (*e.g.*, the dT primer) used in cDNA

library synthesis can incorporate a restriction enzyme recognition sequence. Thus, in some

embodiments, the dT primer used in synthesis of the first "antisense" strand of cDNA

includes a restriction enzyme recognition site, *e.g.*, *Not I*. Thus, the resulting double stranded

5   cDNA would have a restriction enzyme site (*e.g.*, *Not I*) at the site corresponding to the 3'

end of the original sense strand of mRNA. As noted above in connection with modification

of the 5' end of the sense strand, other suitable enzymes are known in the art and can be used

in place of *Not I* (*Not I* is referred to here simply to illustrate the methods by which the

shRNA library can be made). The cDNA can then be digested with an enzyme that

10   recognizes and cleaves the incorporated sequence, and a biotinylated anchoring linker can be

ligated to the 3' end of the paired strands. The biotinylated linker will have an overhanging

sequence containing nucleotides (or "bases"), at least some of which are complementary to

the overhanging bases left on the cDNA after digestion with the selected enzyme (*e.g.*, *Not I*).

See Fig. 1C.

15   <u>Generation of cDNA Tags and Dual Hairpin Structures</u>

To produce cDNA tags from the cDNA library (*e.g.*, templates suitable for use in

making shRNAs), the cDNA bearing one or more anchoring linkers (e.g., biotin at the 5'

and/or 3' ends) is digested with an enzyme anticipated to cut the cDNA with a certain

frequency. For example, one can use an enzyme that cuts the cDNA about once in every

20   256 base pairs on average. This frequency can be achieved with a restriction enzyme that

recognizes a four-base sequence (such enzymes may be referred to as "four-cutters"). Using

a four-cutter, most, if not all, of the cDNAs will be cut at least once. Among the numerous

enzymes that could be used are: *Aci* I, *Alu* I, *Bfa* I, *BfuC* I, *BstU* I, *Cha* I, *Csp6* I, *CviJ* I,

CviR I, *Dpn* I, *Dpn* II, *Fat* I, *Hae* III, *Hha* I, *HinP1* I, *Hpa* II, *HpyCH4* IV, *HpyCH4* V, *Mbo* I,

25   *Mnl* I, *Mse* I, *Msp* I, *Nla* III, *Pho* I, *Rsa* I, *Sau3A* I, *Tai* I, *Taq*$^\alpha$ I, *Tha* I, and *Tsp509* I. Other

suitable enzymes are known in the art (they can be found in scientific publications,

catalogues, and through the worldwide web, including at the website for New England

Biolabs Inc., Beverly, Mass., USA). In one embodiment, the enzyme is *Tai* I. For the sake of

brevity, the following discussion refers to the use of *Tai* I, but any enzyme that cuts, or that is

30   anticipated to cut, cDNA about once every 256 base pairs (*e.g.*, any four-cutter) can be used

in place of *Tai* I. In some embodiments, more than one enzyme is used, concurrently or

sequentially. Where the sequence of a gene (or genes) that is suspected to be of interest is

known, one or more enzymes can be chosen that will cut it at least once within the sequence

of particular interest. In some embodiments, enzymes that cut more or less frequently (*e.g.*,

14

about every 128 or 512 bases) can be used.

        The modified end(s) of the molecule generated to this point can be anchored by way
of the linker (*e.g.*, a biotinylated linker as described above) (this embodiment is illustrated in
Figs. 1B and 1C). After digestion with *Tai* I, the 5' or 3' ends are retained, *e.g.*, with

5   streptavidin magnetic beads (avidin could also be used) that bind to the biotinylated end(s) of
the construct. Thereafter, the biotin linkers are removed from the construct (linkers carrying
BamH I sites can be removed by digestion with *BamH* I; linkers carrying *Not* I sites can be
removed by digestion with *Not* I; and so forth). The linkers can be removed with the
streptavidin beads, leaving double stranded cDNA having one *BamH* I end (and/or one *Not* I

10   end) and one *Tai* I end (as with other intermediates generated in the course of producing an
shRNA library, the digested cDNAs (*e.g.*, *BamH* I and *Tai* I digested cDNAs) are within the
scope of the present invention.

        In some embodiments, the digested duplexes are then linked to a second anchoring
linker. For example, anchoring (*e.g.*, biotinylated) linkers having a type IIS restriction

15   enzyme recognition sequence at the 5' end of the linker ("IIS linkers") can be ligated to the
cut *Tai* I ends (see, *e.g.*, Fig. 1B). Type IIS restriction enzymes cleave at a defined distance
from their recognition site. Although the present discussion refers to the use of *Mme* I (which
in this case cuts 21 bases towards the 5' end of the "antisense" strand), other enzymes that cut
a suitable number of bases distant from or outside the recognition site could also be used. For

20   example, type IIS enzymes include: *Aar* I, *Aci* I, *Alo* I, *Bae* I, *BbvC* I, *Bbv* I, *Bcc* I, *BceA*I,
Bcg I, *BciV* I, *Bfi* I, *Bpu10* I, *Bsa* XI, *BseR* I, *BseY* I, *BsmA* I, *BsmF* I, *Bsm* I, *BspM* I, *BsrB* I,
BsrD I, *Bsr* I, *BstF5* I, *Btr* I, *Bts* I, *Eci* I, *Eco31* I, *Eco57* I, *Eco57M* I, *EcoP15* I, *Esp3* I,
Fau I, *Gsu* I, *Hga* I, *Hph* I, *Ksp632* I, *Mly* I, *Mme* I, *Mnl* I, *Ple* I, *Ppi* I, *Psr* I, *Sap* I, *SfaN* I,
TspDT I, and *TspGW* I. Other suitable enzymes are known in the art, and include those that

25   can be found in the literature or on the internet (*e.g.*, at sites such as neb.rebase.com).

        Typically, the type IIS enzyme will be an enzyme that cuts at least 14 bases away
from its binding or recognition site (such enzymes include *Eco57* I, *Bpm* I, *BspH614* I, *Bco35*
I, *Gsu* I, *Bce83* I, *Bsg* I, *EcoP15* I and *Mme* I (which is used to illustrate the invention). The
construct is then digested with the enzyme selected (in our illustration, *Mme* I), and the

30   construct is purified using a binding partner that recognizes the anchoring moiety on the
linker (*e.g.*, streptavidin beads).

        In some embodiments, illustrated in FIG. 1B, following digestion with the type IIS
enzyme, the first enzyme (*e.g.*, *Tai* I) is used to digest the construct and remove the IIS linker.
The linker can be removed (*i.e.*, purified away from the sample) by virtue of its binding

                                            15

partner (*e.g.*, where a linker includes biotin, it can be removed by binding to streptavidin-coated beads or another streptavidin-coated substrate). The two base pair overhang that results from the *Mme* I digest (represented by NN on figure 1B, 1C, 2A and 2C), can be removed with mung bean nuclease to produce a cDNA tag with a 5' blunt end and a 3' overhang corresponding to the sequence of *Tai* I. When *Mme* I is used, the remaining construct generally comprises about 18 complementary base pairs derived from the sequence of the cDNA; when *Mme* I is used in conjunction with *Tai* I, the construct typically includes 20 base pairs derived from the cDNA.

In another embodiment, the 3' end is anchored, *e.g.*, using a biotinylated linker as described above (this embodiment is illustrated in FIG. 1C). After digestion with *Tai* I, the 3' ends are retained (*e.g.*, with streptavidin magnetic beads) that bind to the biotinylated end of the construct. Thereafter, the biotin linkers are removed from the construct (*e.g.*, by digestion with *Not* I), and purified away with streptavidin beads, freeing double stranded cDNA having, e.g., one *Not* I end and one *Tai* I end (as noted above, the linker can contain a binding moiety other than biotin). New biotinylated linkers having a type IIS restriction enzyme recognition sequence at the 3' end of the linker ("IIS linkers") are ligated to the cut (e.g., *Not* I-cut) 5' ends. In some embodiments, *Mme* I (which in this case cuts 21 bases towards the 5' end of the "sense" strand) is used. In some embodiments, *EcoP15 I* is used.

Although the present discussion refers to the use of *Mme* I, other type IIS restriction enzymes that cut a suitable number of bases away from the recognition site, e.g., as described above and known in the art, could also be used when modifying the 3' end of the cDNA. The construct is digested (*e.g.*, with *Mme* I) and purified using streptavidin beads. When *Mme* I is used, the remaining construct generally comprises about 18 complementary base pairs derived from the sequence of the cDNA; when *Mme* I is used in conjunction with *Tai* I, the construct comprises 20 base pairs derived from the cDNA.

The two base pair overhang resulting from the *Mme* I digest can be removed (*e.g.*, with Mung Bean nuclease), and the biotinylated IIS linker can be removed by digestion with *Not* I, to yield a cDNA tag with a 5' *Not* I end and a 3' blunt end. The biotinylated linker can optionally be left in place to facilitate purification until after addition of a second hairpin, as described herein.

Production of shRNA Expression Constructs

A cDNA tag (produced, for example, by the methods described herein) can be inserted into a vector and expressed as dsRNA. Many appropriate vectors are known in the art. For example, the vector can be a DNA construct (*e.g.*, a plasmid) that is suitable for

16

amplifying and expressing the cDNA tag. The vector can have a sequence that encodes a selectable marker and/or one or more regulatory sequences, such as promoter sequences. For example, the vector can include one or more sequences to which a polymerase will bind and facilitate expression of the inserted template. More specifically, the vector can include two

5   promoters, one oriented on each side of the insert (*i.e.*, the cDNA tag). For example, one can place the cDNA tag between two RNA polymerase promoters (*e.g.*, between two RNA pol III promoters), which will drive transcription towards each other and result in dsRNA species capable of RNAi. Such vectors can also include acceptable transcription initiation and termination sites. This expression vector and populations of such vectors containing cDNA

10 tags made from cDNA libraries by the methods described above, and cells that include them, are within the scope of the present invention. The cDNA library can be made from any of the cell types described above. Accordingly, the invention includes vectors and populations of vectors including cDNA tags made from, for example, mammalian cells (*e.g.*, human cells) that may have been selected (*e.g.*, on the basis of their type, age, or disease phenotype or

15 genotype) or manipulated in some way (*e.g.*, exposed to a therapeutic agent, toxin, metabolite, or the like). Given that the cDNA tags can be prepared, as described above, from cells for which we have incomplete expression data, collections of such cDNA tags are unique (*e.g.*, unique in their representation of essentially all expressed genes) and are within the scope of the present invention.

20   The cDNA tag can also be used to create an shRNA expression template. As the methods of the invention can be used to produce unique cDNA tags, populations of shRNA expression templates produced from those cDNA tags are also unique and are within the scope of the present invention. To create an shRNA expression template, a first hairpin sequence can be added to the 3' or 5' end of the cDNA tag to form a dual hairpin structure.

25 The first hairpin sequence can be any hairpin known in the art, including any hairpin sequence described herein. In one embodiment, the first hairpin sequence is substantially identical to SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 (sequences that are substantially identical to SEQ ID Nos. 1, 2, or 3, include sequences in which one or more nucleotides have been added, deleted, or replaced without preventing the sequence from forming an effective

30 hairpin loop). The first hairpin sequence can be added, e.g., by ligating a pre-formed (*e.g.*, synthetic) hairpin oligonucleotide to the template (*e.g.*, as shown in Figs. 2A-2D). Alternatively, the first hairpin can be added by subcloning the cDNA tag into a vector including the hairpin sequence (*e.g.*, as shown in Figure 3A). A second hairpin can also be added, e.g., as described herein.

17

As a further alternative, a linker, e.g., a non-IIS linker, can include a promoter sequence (*e.g.*, a promoter that can be used to transcribe shRNA); in this case, the linker is retained, and shRNA can be transcribed directly from the shRNA expression template (see Fig. 3B). Individual cDNA tags and/or shRNA expression templates can be isolated by limiting dilution. *In vitro* transcription is achieved by addition of purified RNA polymerase III components, according to methods known in the art. The resulting dsRNA constructs are then purified (purification can also be carried out using methods known in the art) and available for downstream applications.

Generating shRNA Expression Templates

Included herein are a number of methods for generating shRNA expression templates. As one example, a second hairpin, *e.g.*, a hairpin containing two uracil residues in the loop segment (Kaur and Makrigiorgos, *Nucleic Acids Res.* 31:e26, 2003), can be ligated to the dsDNA tag, as shown in Figures 2A-2D. Digestion with uracil glycosylase (or a functionally equivalent enzyme) can then be used to break open the di-uracil loop to allow for synthesis of the second strand. When opened and denatured, e.g., by heating, the dual hairpin structure will consist of a single strand of nucleotides, the sequence of which represents: a first portion of the cleaved loop, a first strand of the dsDNA tag (*e.g.*, the sense or antisense strand), the uncleaved loop, a second strand of the dsDNA tag (*e.g.*, the antisense or sense strand, complementary to the first strand), and the second portion of the cleaved loop. The structure of the construct can be either sense-loop-antisense, or antisense-loop-antisense, where the sense sequence is the same as the corresponding mRNA sequence, and anti-sense is the complement of that sequence. In at least some circumstances, antisense-loop-sense constructs have proven, on average, more efficacious (*see* Khvorova *et al.*, *Cell* 115:209–216, 2003). However, useful hairpin sequences can be generated in either orientation.

When the second, complementary strand is added, the double stranded molecule is referred to herein as an shRNA expression template (*see* the structure at the top of FIGs. 2E and 2F, FIG. 3B). The shRNA expression template can be modified. For example, as described herein, either or both ends can be modified to facilitate cloning (*e.g.*, insertion into a plasmid vector); either or both ends, whether modified or not, can be operably linked to a regulatory sequence (*e.g.*, a promoter); and the sequence of the uncleaved loop can be shortened.

Rather than a uracil glycosylase substrate (*e.g.*, the pair of uracils shown in the drawings), the sequence of one of the hairpins (*e.g.*, the "second" hairpin) can include one or more restriction enzyme recognition sites. In that event, digestion with the corresponding

18

restriction enzyme can be used instead of uracil glycosylase to open the loop. The restriction enzyme can be a "rare" cutter (*e.g.*, an enzyme that recognizes nucleotide sequences at least six, seven, or eight (or more) nucleotides long) to minimize the chance that the template will be cut. As another example, the cDNA tag-hairpin construct can be denatured using methods

5   known in the art, and a second strand can be synthesized to form an shRNA expression template (*e.g.*, as shown in FIG. 3B). Once synthesized, a di-uracil hairpin (or "UU hairpin"), similar to those used in the prior strategies, can be ligated to the template. From here, the library can be created by steps that are the same (or substantially similar to) the steps used in prior strategies.

10   Known techniques can be used to amplify the library. For example, one can grow a number of colonies or plaques containing clones from the library. Large numbers of clones can be grown (*e.g.*, in a multiwell plate), and colonies can be picked manually or robotically (*see, e.g.*, Nguyen *et al.*, *Genomics* 29:207-216, 1995). PCR-based techniques can also be used to amplify the library.

15   Hairpin Sequences

Hairpin sequences suitable for use in the methods described herein can vary in the length of the complementary "stem" and non-complementary "loop" portions. For example, the non-complementary loop portion of the hairpin can range between 4 to 23 nucleotides (*see, e.g.*, Paddison *et al.*, *Genes & Dev.* 16:948-958, 2002). The minimum size of the

20   complementary stem portion is determined by the need for sufficient sequence length to allow for the formation of stable hairpin structures, which facilitates both ligation of pre-formed hairpins onto a cDNA tag, and for self-primed extension, as described above. The length of the hairpin can be reduced by restriction digestion (*e.g.*, after subcloning into a vector as described herein), as is shown in FIGs. 2E, 2F, and 3C; thus, hairpin sequences suitable for

25   use in the present methods can include one or more restriction enzyme recognition sequences. Suitable enzymes are known in the art and can be found on the internet (*e.g.*, at the website of New England Biolabs, Inc., Beverly, Mass. USA). For example, hairpin sequences known in the art can be used (*e.g.*, based on hairpin regions known to those persons skilled in the art to be present in nucleic acid molecules including DNA, tRNA, snRNA, rRNA, mtRNA, or

30   structural RNA sequences). Hairpin structures can be identified by one of skill in the art using, for example, predictive computer modeling programs such as Mfold (Zuker *et al.*, *In RNA Biochemistry and Biotechnology*, pp. 11-43, Barciszewski & Clark, Eds., NATO ASI Series, Kluwer Academic Publishers, 1999); RNAstructure (Mathews *et al.*, *J. Mol. Biol.* 288:911-940, 1999), RNAfold in the Vienna RNA Package (Ivo Hofacker, Institut für

19

Theoretische Chemie, Währingerstr. 17, A-1090 Wien, Austria), Tinoco plot (Tinoco, I. Jr. ,
Uhlenbeck, O. C. and Levine, M. D. Nature 230, 363-367, 1971), ConStruct, which seeks
conserved secondary structures (Steger and Riesner, *J. Mol. Biol.* 258:813-826, 1996; and
Luck *et al.*, *Nucleic Acids Res.* 21:4208- 4217, 1999), FOLDALIGN, (Gorodkin *et al.*,

5   *Nucleic Acids Res.* 25:3724-3732, 1997; and Gorodkin *et al.*, *ISMB* 5:120-123, 1997), and
RNAdraw (Matzura and Wennborg, *Computer Applications in the Biosciences (CABIOS)*,
12:247-249, 1996). Using these programs, hairpin sequences can be identified. Multiple
loop regions can be eliminated to result in shorter, more easily synthesized hairpins.

10   *First Hairpin Sequences*

Suitable first hairpin sequences can be any sequence that forms a hairpin. In some
embodiments, the first hairpin sequence is, or includes, an artificial polynucleotide sequence
capable of forming a stem-loop structure when the polynucleotide is RNA. The first hairpin
sequence can be designed such that the formation of a hairpin results in the creation of an

15   overhanging end compatible for use in ligating the hairpin to the dsDNA-template. In a
preferred embodiment, the first hairpin sequence includes a first region including stem
sequence, a second region including loop sequence, and a third region including stem
sequence, wherein the first and third stem regions comprise regions complementary to each
other, that are at least one nucleotide long. The complementary portions of the first and third

20   regions do not need to be, and typically are not, very long, and can be as short as one or two
nucleotides. The first hairpin sequence can also include one or more restriction enzyme
recognition sites placed such that digestion with the restriction enzyme removes part of the
hairpin sequence, leaving a short loop of up to about 6 to about 10 nucleotides. Typically, the
melting temperature of the hairpin will be between 70°C and 75°C and have a GC content

25   less than 75%. For example, the first hairpin sequence can be one of the sequences shown in
Table 1. Digestion with *Bcg* I can be used to remove much of the sequence, leaving a short
loop connecting the two regions of the dsDNA complementary to the original cDNA.

20

Exemplary First Hairpin Sequences

Use with	Sequence	SEQ ID NO:
Dpn II	GATC*TCCAAGAGA*|*ATAGCAGCT*C<u>*CGATATCGCTGCT*</u>*ATTCTCTTGGA*|	1
Tai I	*TCCAAGAGA*|*ATAGCAGCT*C<u>*CGATATCGCTGCT*</u>*ATTCTCTTGGA*|	2
Tsp509 I	*AATTTCCAAGAGA*|*ATAGCAGCT*C<u>*CGATATCGCTGCT*</u>*ATTCTCTTGGA*|	3

Underlined regions denote Bcg I restriction enzyme recognition sites. Pipes indicate where Bcg I cuts. Italics indicate complementary stem regions.

5   Suitable RNA stem-loop structures can be found in databases such as the Small RNA

Database (Perumal *et al.*, Department of Pharmacology, Baylor College of Medicine, USA),

Database of non-coding RNAs (Erdman *et al.*, *Nucleic Acids Res.* 29:189-193, 2001)), large

subunit rRNA database (Wuyts *et al.*, *Nucleic Acids Res.* 29:175-177, 2001), the small

subunit rRNA database (Wuyts *et al.*, *Nucleic Acids Res.* 30:183-185, 2002), snoRNA

10   Database for budding yeast (Lowe and Eddy, *Science* 283:1168-1171, 1999) for *Archaea*

(Omer *et al.*, *Science* 288:517-522, 2000), for *Arabidopsis thaliana* (Brown *et al.*, *RNA*

7:1817-1832, 2001), tRNA sequences and sequences of tRNA genes (Mathias *et al.* on the

world wide web at uni-bayreuth.de/departments/biochemie/trna/), the 5S ribosomal RNA

database (Szymanski *et al.*, *Nucleic Acids Res.* 30:176-178, 2002), The Nucleic Acid

15   Database Project (NDB) at Rutgers University (on the world wide web at

ndbserver.rutgers.edu/NDB/), and The RNA Structure Database (on the world wide web at

RNABase.org).

          In some embodiments, the first hairpin sequence is derived, *e.g.*, from an artificially

generated polynucleotide sequence predicted to form a hairpin. In other embodiments, the

20   hairpin sequence is derived from hairpin regions substantially identical to a lin-4 or let- 7

homolog sequence found in the host cell or a portion thereof (*e.g.*, the human let-7 homolog,

miR-98), or some other naturally occurring hairpin sequence or a portion thereof.

          Second Hairpin Sequences

          Suitable second hairpin sequences can be any sequence predicted to form a stable

25   hairpin. The second hairpin sequence can include a specific sequence that allows the hairpin

to be broken open, to form a double-stranded structure with non-complementary ends. As

one example, the specific sequence can be two uracil residues, and treatment with uracil

glycosylase followed by heating can be used to break the second hairpin open. As another

example, the specific sequence can be a recognition site for a restriction enzyme. A

30   restriction enzyme with a relatively rare recognition sequence, *e.g.*, at least a six-, seven- or

21

eight-cutter (*i.e.*, a restriction enzyme having a six, seven, or eight base pair recognition sequence), can be used to lessen the likelihood that the dsDNA-loop template will be cut; some six-cutters might cut too frequently to be useful in most plants and animals, but six-cutters would be useful for prokaryotes and those eukaryotes with small genomes; six-cutters with all GCs in their restriction sites can also be used for eukaryotes with normal-sized genomes that are AT-rich. Suitable enzymes are known in the art and can be found on the internet, *e.g.*, at neb.rebase.com. The second hairpin sequence can also contain one or more restriction enzyme recognition sites for use in subcloning into a vector. In one example, the second hairpin sequence can be: CGAAGAGCGCCTGCTTGAGATGCTGT TGAG*ACGT*CGUUACTATCCTTGAACAGCGCTCTTCG (SEQ ID NO:4); as shown here, the two uracil residues are in bold face, and an *HpyCH4* IV restriction enzyme recognition site is underlined. In other embodiments, the second hairpin sequence is as follows: C<u>GGATCC</u>ATTCCGGG*TCCCG*CTGCTGGCGCGUUAGACCGGCCGCCGTCAGCCGCCA TCGGCCAAT<u>GGATCC</u>GACGT (SEQ ID NO:5); here, the two uracil residues are in bold face, a *BamH* I restriction enzyme recognition site (in the double-stranded stem portion) is underlined; and a *BsmF* I restriction enzyme recognition site is in italics.

One of skill in the art will appreciate that there are a large number of sequences suitable for use as second hairpin sequences. Preferred sequences comprise 1) a pair of UU-nucleotides; 2) a type IIs restriction enzyme recognition site, *e.g.*, a *BsmF* I site; and 3) an 8-10 base pair stem with a large loop, *e.g.*, at least 20 base pairs long. This provides room for the primers to bind. Typically, the melting temperature for the hairpin and the primers will be similar.

Primers can be designed that are complementary to the sequence of the second hairpin, and once the second hairpin is broken open (*e.g.*, using uracil glycosylase or restriction digest as described herein) the primers can be used to synthesize a second full strand to make an shRNA expression template construct having two identical dsDNA-template regions separated by the first hairpin sequence, substantially as shown in Figure 2E. As one example, where the second hairpin has the sequence of SEQ ID NO:4, primers can be used with one of the sequences shown in the following table:

Primer	Sequence	SEQ ID NO:
UU-Anti	CTCAACAGCATCTCAAGCAG	6
UU-loop-anti 10	CTCAACAGCATCTCAAGCAGGCGCTCTTCG	7
UU-sense	ACGTCGACTATCCTTGAACAG	8
UU-sense-10	ACGTCGACTATCCTTGAACAGCGCTCTTCG	9

22

Where the second hairpin has the sequence of SEQ ID NO:5, primers having the sequences shown in the following table can be used:

Primer	Sequence	SEQ ID NO:
UU-Sense	ACCGGCCGCGTCAGCCGCCATCGGCC	10
UU-anti-sense	CGGATCCATTCCGGGTCCCGCTGCTGGCGC	11

Where the second hairpin has been designed with a restriction enzyme site suitable for subcloning, the resulting dual-template constructs can be digested with that restriction enzyme for insertion of the dual-template construct into a vector. Where the second hairpin has the sequence of SEQ ID NO:4, digestion with the *Ear* I restriction enzyme produces an overhang suitable for insertion into a *BsmF* I site, such as is present on a number of commercially available vectors (FIG. 3C).

Where the second hairpin has the sequence of SEQ ID NO:5, digestion with *BamH* I produces overhangs suitable for insertion into a *BamH* I site. Alternatively, where the second hairpin has the sequence of SEQ ID NO:5, the construct can be digested with *BsmF* I, then the *BsmF* I-cut end can be blunted (*e.g.*, the overhang can be removed, *e.g.*, using mung bean nuclease), then *BamH* I can be used to cut the other end, producing a construct suitable for directional cloning (see FIG. 2E).

Vectors

Numerous vectors suitable for use in the present methods are known in the art or could be constructed by one of skill in the art. The vector can have one or more multiple cloning regions containing a number of restriction enzyme sites for facilitating cloning into the vector. Expression vectors can also contain one or more polymerase promoter sites. Since the dual-template construct is bi-directional, a vector having multiple RNA polymerase promoter sequences can be used. Suitable RNA polymerases include RNA pol III and pol II promoters, including but not limited to U6 and H1. The vector can also contain one or more of the following: a reporter gene (*e.g.*, eGFP, eCFP, β-gal); a T5 termination signal (5 thymidines); other promoters (*e.g.*, SV40; bla); positive or negative selection genes (*e.g.*, antibiotic resistance genes, *e.g.*, neomycin-R; thymidine kinase), and an origin of replication, (*e.g.* f1, LTRs). The vector can be viral or non-viral, *e.g.*, a plasmid. Viral vectors can include, *e.g.*, adenovirus, adeno-associated virus, and retrovirus, *e.g.*, as known in the art, *e.g.*, as described in Brummelkamp *et al.*, *Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer Cell.* 2(3):243-7, 2002. Inducible promoters can also be

23

used, *e.g.*, as known in the art, *e.g.*, as described in Van De Wetering *et al.* (*EMBO Rep.* 4:609-615, 2003). A number of commercially available vectors are available and can be used or modified and used for use in the present methods. Such vectors include, but are not limited to: pSilencer™ (Ambion); pSuper (Brummelkamp *et al.*, *Science* 296:550-553, 2002); psiRNA™ (Invivogen); and pSuppressor™ and pSuppressorAdeno™ (Imgenex).

Methods of Use

Any of the nucleic acids described herein can be introduced into a biological cell or population of cells (whether clonal or diverse). For example, an shRNA, shRNA library, or a subset thereof, can be chemically synthesized or made (*e.g.*, transcribed from a DNA template) *in vitro* or in cell culture and then introduced into a cell by any of the art-recognized methods for transducing (*e.g.*, transfecting) cells (*e.g.*, introduced by calcium phosphate precipitation, lipofection, or biolistics). Alternatively, an expression vector (*e.g.*, a plasmid or viral vector) that includes a sequence encoding a nucleic acid of interest (*e.g.*, an shRNA, shRNA library, or a subset thereof), e.g., as shown in FIGs. 2E, 2F, and 3C, can be introduced into a cell. The subsequent expression can be transient or stable.

The nucleic acid libraries of the present invention have a number of uses, including the identification of sequences as therapeutic agents or as targets for therapeutic agents. The use of normalized and/or subtracted cDNA libraries as a starting material allows more effective screening of low-abundance and differentially expressed genes, which may be of particular interest. No prior knowledge of gene sequences or expression patterns is necessary, as the methods can be used to generate a library of shRNAs and/or shRNA expression templates from a pool of unknown cDNAs (*i.e.*, cDNAs generated against mRNA sequences that are incompletely defined or identified).

An shRNA library, or a fraction thereof, produced by the methods described herein can be screened by introducing the shRNA(s) into a cell or population of cells (*e.g.*, cells in culture or *in vivo* (*e.g.*, cells within a simple organism)) to identify potential therapeutic targets or compounds. As large numbers of cells can be cultured (*e.g.*, in one or more multi-well plates) and screened at essentially the same time, the screening methods can be configured as high-throughput screens. *See, e.g.*, Ziauddin and Sabatini, *Nature* 411:107-110, 2001, and Wu *et al.*, *Trends Cell. Biol.* 12:485-488, 2002, which are incorporated herein by reference. For example, the shRNAs or shRNA clones produced by the methods described herein can be introduced, singly or in pools, into cells or organisms, and a preselected parameter of the cells or organisms can be investigated (the "parameter" is varied and is described further herein).

24

Suitable cells include both primary cells and cells that have been modified (*e.g.*, immortalized) and/or passaged in tissue culture. Where the screen is conducted *in vivo*, the organism can include any living organism, whether plant or animal. "Simple" organisms can be used in the initial stages of identifying therapeutic genes and include, but are not limited

5   to, *C. elegans, D. melanogaster,* and *D. rerio*, as well as bacteria and other prokaryotes (*e.g.*, fungi and yeast).

Once transfected with an shRNA, the parameter studied can be any measurable parameter associated with the cell or organism. For example, the parameter can be a detectable change in the gross state of an organism, such as the state of a disease (*e.g.*,

10   regression of a cancer; a change in neural activity, behavior, or mental capacity (*e.g.*, improved memory or other cognitive skill); or a change in physical ability (*e.g.*, improved balance or ability to move)). The parameter can also be a more particular biochemical activity (*e.g.*, enzyme activity) as compared to a reference. The parameter can be manifest as a change in cellular phenotype (*e.g.*, morphology, proliferation, movement, development,

15   viability or death). Where the parameter involves enzyme activity, the enzyme can be, for example, a kinase (*e.g.*, a MAP kinase), phosphatase, protease, helicase, or polymerase, and may be ATP-dependent or ATP-independent. Other parameters include ion concentrations, fluxes, or gradients (*e.g.*, plasma membrane potential, mitochondrial potential, calcium concentrations) and ligand-receptor binding and/or activation. Other parameters (*e.g.*,

20   metabolic, pathophysiological or developmental parameters), can be also be monitored and evaluated.

As noted, high-throughput screens can be used in the present methods, and those screens can be configured in the same manner as others conducted previously with agents other than shRNAs. In most instances, high-throughput screens are designed to identify

25   agents (here, shRNAs) that affect a selected parameter (*see, e.g.*, Walters and Namchuk, *Nat. Rev. Drug Discov.* 2:259-266, 2003). A parameter on the cells can be any increase, decrease, or other modulation or alteration of the parameter, and a parameter that reaches a certain threshold can be considered a positive result. For example, a positive result can be assigned when a parameter, or a change in the parameter, reaches a predetermined level of modulation

30   (*e.g.*, inhibition or activation). A positive result can alternatively be assigned by defining a point at which the parameter, or change in the parameter, reaches statistical significance. In a third method, a number of positive results can be defined that will be followed up, *e.g.*, the thousand, hundred, or ten compounds that cause the greatest effect on the cells or organisms, *e.g.*, the largest change in one or more parameters of the cells or organisms.

25

In one embodiment, the parameter that is investigated is relevant to a pathophysiological state, *e.g.*, a disease state. The disease state can be, e.g., any cell-autonomous pathology, *e.g.*, any pathology that is a candidate for library based screens or a potential target for RNAi based therapies.

5   For example, the parameter can be related to a disorder associated with unwanted or undesirable cellular proliferation or differentiation, *e.g.*, cancers or skin disorders. Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, *e.g.*, leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited 10   to those of prostate, colon, lung, breast and liver origin.

As used herein, the terms "cancer," "hyperproliferative," and "neoplastic" refer to cells having the capacity for autonomous growth, *i.e.*, an abnormal state or condition characterized by rapidly proliferating cell growth. Hyperproliferative and neoplastic disease states may be categorized as pathologic, *i.e.*, characterizing or constituting a disease state, or 15   may be categorized as non-pathologic, *i.e.*, a deviation from normal but not associated with a disease state. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. "Pathologic hyperproliferative" cells occur in disease states characterized by malignant tumor growth. Examples of non-pathologic 20   hyperproliferative cells include proliferation of cells associated with wound repair.

The terms "cancer" or "neoplasms" include malignancies of the various organ systems, such as affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of 25   the lung, cancer of the small intestine and cancer of the esophagus.

The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. In some embodiments, 30   the disease is renal carcinoma or melanoma. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, *e.g.*, which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures.

26

The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

Additional examples of proliferative disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, *e.g.*, arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, *e.g.*, erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, *Crit Rev. in Oncol./Hemotol.* <u>11</u>:267-297, 1991); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

Other examples of proliferative and/or differentiative disorders include skin disorders. The skin disorder may involve the aberrant activity of a cell or a group of cells or layers in the dermal, epidermal, or hypodermal layer, or an abnormality in the dermal-epidermal junction. For example, the skin disorder may involve aberrant activity of keratinocytes (*e.g.*, hyperproliferative basal and immediately suprabasal keratinocytes), melanocytes, Langerhans cells, Merkel cells, immune cell, and other cells found in one or more of the epidermal layers, *e.g.*, the stratum basale (stratum germinativum), stratum spinosum, stratum granulosum, stratum lucidum or stratum corneum. In other embodiments, the disorder may involve aberrant activity of a dermal cell, *e.g.*, a dermal endothelial, fibroblast, immune cell (*e.g.*, mast cell or macrophage) found in a dermal layer, *e.g.*, the papillary layer or the reticular layer.

Examples of skin disorders include psoriasis, psoriatic arthritis, dermatitis (eczema), *e.g.*, exfoliative dermatitis or atopic dermatitis, pityriasis rubra pilaris, pityriasis rosacea, parapsoriasis, pityriasis lichenoiders, lichen planus, lichen nitidus, ichthyosiform dermatosis, keratodermas, dermatosis, alopecia areata, pyoderma gangrenosum, vitiligo, pemphigoid (*e.g.*, ocular cicatricial pemphigoid or bullous pemphigoid), urticaria, prokeratosis, rheumatoid arthritis that involves hyperproliferation and inflammation of epithelial-related

27

cells lining the joint capsule; dermatitises such as seborrheic dermatitis and solar dermatitis; keratoses such as seborrheic keratosis, senile keratosis, actinic keratosis. photo-induced keratosis, and keratosis follicularis; acne vulgaris; keloids and prophylaxis against keloid formation; nevi; warts including verruca, condyloma or condyloma acuminatum, and human

5 papilloma viral (HPV) infections such as venereal warts; leukoplakia; lichen planus; and keratitis. The skin disorder can be dermatitis, *e.g.*, atopic dermatitis or allergic dermatitis, or psoriasis.

In some embodiments, the disorder is psoriasis. The term "psoriasis" is intended to have its medical meaning, namely, a disease which afflicts primarily the skin and produces

10 raised, thickened, scaling, nonscarring lesions. The lesions are usually sharply demarcated erythematous papules covered with overlapping shiny scales. The scales are typically silvery or slightly opalescent. Involvement of the nails frequently occurs resulting in pitting, separation of the nail, thickening and discoloration. Psoriasis is sometimes associated with arthritis, and it may be crippling. Hyperproliferation of keratinocytes is a key feature of

15 psoriatic epidermal hyperplasia along with epidermal inflammation and reduced differentiation of keratinocytes. Multiple mechanisms have been invoked to explain the keratinocyte hyperproliferation that characterizes psoriasis. Disordered cellular immunity has also been implicated in the pathogenesis of psoriasis. Examples of psoriatic disorders include chronic stationary psoriasis, psoriasis vulgaris, eruptive (gluttate) psoriasis, psoriatic

20 erythroderma, generalized pustular psoriasis (Von Zumbusch), annular pustular psoriasis, and localized pustular psoriasis.

The parameter can also be related to a nervous system disorder, *e.g.*, a neurodegenerative disorder, *e.g.*, aceruloplasminemia, adrenoleukodystrophy, Alzheimer's disease, amyotrophic lateral sclerosis, Angelman syndrome, ataxia telangiectasia,

25 CharcotMarieTooth syndrome, Cockayne syndrome, Creutzfeldt-Jakob disease, deafness, Duchenne muscular dystrophy, epilepsy, essential tremor, familial mediterranean fever, fragile X syndrome, Friedreich's ataxia, Gaucher disease, Huntington's disease, Machado-Joseph disease (Spinocerebellar Ataxia Type 3), maple syrup urine disease, Menkes syndrome, myotonic dystrophy, neurofibromatosis, Niemann-Pick disease, Parkinson's

30 disease, phenylketonuria, Prader-Willi syndrome, Refsum disease, Rett syndrome, spinal muscular atrophy, spinocerebellar ataxia, Tangier disease, Tay-Sachs disease, tuberous sclerosis, Von Hippel-Lindau syndrome, Williams syndrome, Wilson's disease, and/or Zellweger syndrome. In some embodiments, the neurodegenerative disorder can be an inherited neurodegenerative disorder, *e.g.*, aceruloplasminemia; a polyglutamine expansion

28

disease including any of several spinocerebellar ataxias and Huntington's disease; Parkinson's disease; and/or familial amyotrophic lateral sclerosis. In some embodiments, a model of a disease can be used in the screen, including tissue culture or primary cell models, or simple organismal models example, a cell model of a neurodegenerative disease can be used, and an effect on a parameter relevant to the neurodegenerative disease would be an indicator of a positive result. shRNA clones identified as positive results may have therapeutic activity, or may be indicators of a useful target for therapeutic intervention. "Therapeutic activity" can include activity that is useful to treat, delay, or prevent the development or progression of a disease. shRNA clones demonstrated to have therapeutic activity may be useful as therapeutic compounds.

The screen can further include the presence and/or absence of one or more non-shRNA compounds, for example, therapeutic compounds (compounds known to have therapeutic activity) or candidate therapeutic compounds (compounds suspected to have therapeutic activity).

Identified clones can be further screened in one or more secondary screens (*e.g.*, in an animal model) such as a non-human mammal. Clones that pass the secondary screens (by, for example, correcting or ameliorating the disease phenotype) are potential therapeutic reagents. Such clones can be placed into systems for the delivery of DNA or RNA into a subject (*e.g.*, an experimental animal or a human). If the target of the shRNA clone already has known inhibitors, they could be used in conjunction with each other to bolster the therapeutic efficacy.

The invention is further described in the following examples, which do not limit the scope of the invention claimed.

EXAMPLES

Example 1. Preparation of a normalized cDNA library

Normalized cDNA libraries can be produced by a technique adapted from Carninci *et al.* (*Genome Res.* 10:1617-1630, 2000).

5   *RNA isolation*: Cells are harvested on ice in phosphate buffered saline (PBS) and lysed with lysis buffer (100 mM NaCl, 5 mM MgCl$_2$, 50 mM Tris (pH 7.5), 0.5% NP-40, 10 mM vanadyl ribonucleoside, 5000 units RNase inhibitor). The RNA is precipitated with cetyltriammonium bromide (CTAB) and urea. After resuspension in 7 M guanidinium chloride, the RNA is purified by phenol:chloroform, chloroform extraction, and poly(A)$^+$

10   RNA is isolated from total RNA with the PolyA-Quick™ kit from Stratagene.

First-strand synthesis of cDNA: The poly(A)$^+$ mRNA is mixed with reverse transcriptase. In addition to the provided reaction buffer, dithiothreitol (DTT), 10 mM dNTPs, sorbitol, trehalose, and the primer adapter are added. Unlike the original protocol of Carninci *et al.*, 5-methyl-dCTP was not included, as cleavage of the resulting

15   cDNA is required. The sequence of the degenerate primer adapter is

Degenerate primer adapter	5'-GAGAGAGAGAAAGGATCCTTTTTTTTTTTTTTTTTTTVN*-3'	SEQ ID NO:12

*where V is G, A, or C, and N is G, A, T, or C. A *BamH* I site is underlined.

Synthesis is performed in a thermal cycler programmed as follows: 40°C, for 4

20   minutes; 50°C for 2 minutes; and 56°C for 60 minutes. The reaction can be quantified by running a tube containing [α-^{32}P-dGTP] in parallel. The resulting cDNA/mRNA hybrids are cleaned by proteinase K treatment and precipitated with CTAB/urea.

Full-length cDNA recovery: To purify full-length cDNA/mRNA hybrids, the cap-structure at the 5' termini and the 3' termini are oxidized and subsequently biotinylated.

25   Oxidation is accomplished by incubating hybrids with 10 mM NaIO$_4$ for 45 minutes. Following isopropanol precipitation, the hybrids are incubated overnight with 10 mM biotin hydrazide long arm. The biotinylated hybrid is precipitated with sodium citrate, NaCl, and ethanol. After resuspension, cDNA/mRNA hybrids that are biotinylated but that do not contain two full-length strands are eliminated by RNase I digestion. Full-length

30   cDNA/mRNAs can then be recovered by using magnetic streptavidin beads. The nucleic acids are eluted with 50 mM NaOH and 5 mM EDTA. This solution also denatures the hybrids.

Second strand synthesis: The eluted single-strand cDNA is treated with RNase I to ensure the complete removal of all RNA. To remove traces of the primer used in first-strand synthesis, the single-strand cDNA is passed through an equilibrated s400 spun column (Amersham Biosciences). The Single-Strand Linker Ligation method (Shibata *et al.*,

5   *Biotechniques* <u>30</u>:1250-1254, 2001) is adapted to prime second strand synthesis. The oligonucleotides originally published are modified to be 5' biotinylated and to include a *Not* I restriction site. Annealed oligonucleotides are added to the single-strand cDNA and ligated overnight. The reaction is stopped with EDTA and SDS. After phenol:chloroform, chloroform extraction, the excess linker is eliminated with S300 spun column

10   chromatography.

Normalization of cDNA: Normalization (a process whereby one attempts to obtain a more equal number of cDNAs corresponding to messages originally present in varying amounts) is achieved by mixing the single stranded cDNA with biotinylated mRNA derived from the original sample under conditions in which frequently occurring species will

15   hybridize while rare species will remain unbound. First, the mRNA is biotinylated with the Mirus Label-IT™ nucleic acid biotinylation kit (Panvera). After biotinylation with the kit, the mRNA is ethanol precipitated. After resuspension, biotin-mRNA is mixed with the single-stranded cDNAs at an RoT of 10 in a mix of 80% formamide, 250 mM NaCl, 25 mM HEPES (pH 7.5), and 5 mM EDTA. RoT is the ratio at which half of the RNA will hybridize

20   under set conditions. In this case, an RoT of 10 should provide conditions in which the frequent species have hybridized, but the rare ones remain unbound. An RoT $10 = 0.97$ ug/ul of RNA for 1 hr. If the RNA concentration is doubled, the time to reach an RoT of 10 is halved.

After hybridization, the cDNA/mRNA hybrids are ethanol precipitated. After

25   resuspension, imprecise hybrids are eliminated by RNase I treatment, followed by phenol:chloroform, chloroform extraction. Normalization is achieved by mixing the hybrids with magnetic streptavidin beads. Those cDNAs not hybridized remain in solution, while those bound to biotin-mRNA are removed. The supernatant is concentrated with Microcon-100™ (Millipore) ultrafiltration according to the manufacturer. The resulting cDNA is treated

30   with RNase to remove all traces of RNA and filtered again with a Microcon-100™.

Second-Strand cDNA synthesis: To generate the second strand of cDNA, a second-strand primer (see SEQ ID NO: 13, referred to on the figures as NotI-GA) is mixed with the first strand of cDNA, dNTPs, Elongase™ (Invitrogen), and reaction buffer.

31

2ⁿᵈ strand primer	5'-AGAGAGAGAGAGGCGGCCGCCTCATTTAGGTGACACTATAGAACCA-3' The *Not* I site is underlined.	SEQ ID NO:13

The mix is heated with the following program: 65°C for 5 minutes; 68°C for 30

minutes; and 72°C for 10 minutes. The product is subsequently ethanol precipitated and

5   resuspended in TE. Selected steps of the process described above are illustrated in the

schematic diagram of FIG. 4; the starting and end points are illustrated in FIG. 1A.

Example 2. cDNA tags From the 5' End of the Sense Strand of a Normalized cDNA Library

    Creation of cDNA tags: The cDNA generated at the end of Example 1 is digested

10   with *BamH* I. Biotinylated *BamH* I linkers, having the nucleic acid sequence shown below,

are annealed and ligated to the *BamH* I-digested cDNA in an overnight reaction with T4

ligase. The sequence that anneals to the cDNA to reconstitute the *BamH* I site is underlined

in SEQ ID NO:15.

Sense	5'-AGAGAGAGAGGCGGCCGCCTCATTTAGGTGACACTATAGAAG-3'	SEQ ID NO:14
Antisense	5'-<u>GATCC</u>TTCTATAGTGTCACCTAAATGAGGCGGCCGCCTCTCTCTCT-3'	SEQ ID NO:15

15   Excess linkers are removed by running the cDNA over an s300 spun column

(Amersham Biosciences). The resulting cDNA is digested with *Tai* I. The biotinylated ends

are bound by streptavidin-coated magnetic beads, which are then washed 5X with 4.5 M

NaCl and 50 mM EDTA at pH 8.0. To release the cDNA from the beads, it is exposed to

20   *BamH* I, which releases *BamH* I-*Tai* I tags into the supernatant. The beads are washed 1X

with 1 M NaCl, 10 mM EDTA. The pooled eluates are concentrated with a Microcon-30™

filter. Biotinylated *Mme* I-*Tai* I linkers are annealed and ligated in an overnight reaction to

the *BamH* I-*Tai* I fragments eluted from the column. The sequence that anneals to the cDNA

fragments to reconstitute the *Tai* I site is underlined in SEQ ID NO:16.

Sense	5'-GTAATACGACTCACTATAGGGCGGATCCTCCA<u>ACGT</u>-3'	SEQ ID NO:16
Antisense	5'-TGGAGGATCCGCCCTATAGTGAGTCGTATTAC-3'	SEQ ID NO:17

25   The larger, ligated fragments are then digested with *Mme* I, which digests the cDNA

at a certain distance away from the *Mme* I recognition site and thereby frees the 5' end from

the biotinylated 3' end. The *Mme* I linker and the cDNA that remains bound thereto are

subsequently digested with *Tai* I, which frees an 18 bp cDNA, which is referred to herein as a

30   cDNA tag. The *Mme* I linkers are separated from the cDNA tags by magnetic streptavidin

32

beads. The second digest with *Tai* I, which frees the cDNA tag, is shown at the conclusion of FIG. 1B and the cDNA tag is shown at the beginning of FIG. 2A).

Example 3. cDNA tags From the 3' End of the Sense Strand of a Normalized cDNA Library

The normalized cDNA generated in Example 1 is digested with *Not I* and biotinylated *Not I* linkers are added (a partial *Not* I site is underlined in SEQ ID NO:18).

| Sense | 5'-GG<u>CCGC</u>GAGAGAGAGAAAGGATCC-3' | SEQ ID NO:18 |
| Antisense | 5'-GGATCCTTTCTCTCTCTCG-3' | SEQ ID NO:19 |

The linkers are annealed and subsequently ligated by overnight incubation with T4 ligase. Excess linkers are removed by running the cDNA over an s300 spun column (Amersham Biosciences). The resulting cDNA is digested with *Tai* I. The 3' ends are recovered by their affinity for streptavidin-coated magnetic beads. The beads are washed 5X with 4.5 M NaCl, 50 mM EDTA, pH 8.0. The cDNA is digested with *Not* I, which releases *Not* I-*Tai* I tags into the supernatant. The beads are washed 1X with 1 M NaCl, 10 mM EDTA. The pooled eluates are concentrated with a Microcon-30™ filter. Biotinylated *Mme* I-*Tai* I linkers are annealed and ligated with *Not* I-*Tai* I fragments in an overnight reaction (a partial *Tai* I site is underlined in SEQ ID NO:20).

| Sense | 5'-GTAATACGACTCACTATAGGGCGGATCCTCCA<u>ACGT</u>-3' | SEQ ID NO:20 |
| Antisense | 5'-TGGAGGATCCGCCCTATAGTGAGTCGTATTAC-3' | SEQ ID NO:21 |

The ligated fragments are digested with *Mme* I, which frees the 3' end of the sense strand. The biotinylated *Mme* I linker and cDNA tag are retained on streptavidin beads, and a subsequent digest with *Tai* I frees the 18 bp cDNA tags. The second digest with Tai I, which frees the cDNA tag, is shown at the conclusion of FIG. 1C and the cDNA tag is shown at the beginning of FIG. 2B.

Example 4. Dual Hairpin Amplification

Hairpin addition and second strand synthesis: Two hairpins are added to the two ends of the cDNA tags (as shown in FIGs. 2A and 2C for cDNA tags generated from the 5' end of the sense or antisense strand, respectively, and as shown in FIGs. 2B and 2D for cDNA tags generated from the 3' end of the sense or antisense strand, respectively). Either loop can be added first. Directional addition is guaranteed by the *Tai* I overhang (5'-ACGT-3') that remains on the tags, so the loop that is going to added to this end will typically be added first, then the construct is treated with a nuclease to blunt the ends

33

(removing the 2 bp overhang left from the MmeI digestion) for the addition of the second loop. The loops are boiled for 3 minutes at 97°C and cooled to 37°C over a 30-minute period prior to the ligation.

Tai I loop	5'- TCCAAGAGAATAGCAGCTCCGATATCGCTGCTATTCTTTGGA-3'	SEQ ID NO:2
UU loop	5'-CGGATCCATTCCGGGTCCCGCTGCTGGCGCGUUAGACCGGCCG-CGTCAGCCGCCATCGGCCAATGGATCCG<u>ACGT</u>-3'	SEQ ID NO:5

5   The UU loop can have an ACGT overhang, and thereby anneal with the *Tai* I overhang, as shown above and in FIG 2B, but it can also be blunt and anneal to the non-overhanging side of the cDNA tag (as shown in FIG. 2A). Similarly, the *Tai* I loop may be blunt ended and ligated to the blunt end of the cDNA tag (as shown above and in FIG. 2B),

10   but the *Tai* I loop may also have a compatible *Tai* I overhang (as shown in FIG. 2A).
        The ligation is terminated by proteinase K treatment, followed by phenol:chloroform and chloroform extractions. The cDNA is then digested with *Sac* II to eliminate UU-loop dimers. It is preferable to eliminate the UU-loop dimers, as the primer for second strand synthesis binds to the UU-loop. The portion of the UU-loop that is excised is eliminated by

15   column chromatography. Double hairpin-containing constructs are then linearized with uracil glycosylase, and the second strand is synthesized by mixing the UU-second-strand primer (5'- CGGATCCATTCCGGGTCCCGCTGCTGGCGC-3' (SEQ ID NO:11)) with cDNA, dNTPs, Titanium Taq™ (Invitrogen), and reaction buffer, and heating the construct to denature the loops. The mix is heated with the following program (65°C for 5 minutes; 68°C

20   for 30 minutes; and 72°C for 10 minutes). The product is subsequently ethanol precipitated and resuspended in TE.
        Exemplary schemes for ligating the two hairpins to the cDNA tags are shown in FIGs. 2A-2D. FIG. 2A illustrates a procedure for use in producing a sense-loop-antisense structure from a cDNA tag from the 5' end of the sense strand. FIG. 2B illustrates a

25   procedure for use in producing a sense-loop-antisense structure from a cDNA tag from the 3' end of the sense strand. FIG. 2C illustrates a procedure for use in producing an antisense-loop-sense structure from a cDNA tag from the 5' end of the sense strand. FIG. 2D illustrates a procedure for use in producing an antisense-loop-sense structure from a cDNA tag from the 3' end of the sense strand.

30   Once the second strand is synthesized, an shRNA expression template results, as shown at the top of FIGs. 2E (which shows a sense-loop-antisense structure) and 2F (which shows an antisense-loop-sense structure).

34

Example 5. The Production of shRNA from shRNA Expression Templates

    Placement into a vector and elimination of extraneous sequences: The resulting double-stranded cDNA – the shRNA expression template -- is digested with *BsmF* I, blunt ended with mung bean nuclease, and digested with *BamH* I. The digested template is cloned into a pH1 plasmid vector, which is derived from pBluescript™, that has also been digested with *BamH* I and blunted *BsmF* I. The pH1 plasmid includes an H1 promoter. After overnight ligation with T4 ligase, the cDNA is proteinase K treated and phenol:chloroform and chloroform extracted. The insert within the plasmid is subsequently digested with *Bcg* I to remove extraneous sequences from the loop region and recircularized in a ligation reaction overnight. The resulting plasmid is used to transfect TOP10 bacteria (Invitrogen), which express the insert as shRNA.

Example 6. *Tai* I-Derived shRNA

    Oligonucleotides encoding shRNAs against enhanced green fluorescent protein (EGFP) were obtained from Proligo (Boulder, CO). pAntiGFP was based on published sequences known to silence EGFP.

Sense	5'-GGCTACGTCCAGGAGCGCATTCAAGAGATGCGCTCCTGGACGTAG CCTTTTTT-3'	SEQ ID NO:22
Antisense	5'AATTAAAAAAGGCTACGTCCAGGAGCGCATCTCTTGAATGCGCTC CTGGACGTAGCCGGCC-3'	SEQ ID NO:23

    pTaiI was homologous to the sequence immediately 3' of the *Tai* I site in the cDNA of EGFP.

Sense	5'-CGTCTATATCATGGCCGACTTCAAGAGAGTCGGCCATGATATAG-ACGTTTTTT-3'	SEQ ID NO:24
Antisense	5'-AATTAAAAAACGTCTATATCATGGCCGACTCTCTTGAAGTCGGCC-ATGATATAGACGGGCC-3'	SEQ ID NO:25

    Oligos were annealed and placed into the pSilencer™ vector (Ambion) according to the manufacturer's protocol. The vector p65Q consisted of coding for 65 glutamines followed by eGFP. All vectors were sequence verified. HeLa cells were transfected with p65Q and pAntiGFP, pTaiI, or empty pSilencer™. Transfections were completed with Lipofectin™ and PLUS™ reagent (Invitrogen). After 48 hrs, cells were lysed with RIPA. Fluorescence was measured on a Victor^2 V™ plate reader (Perkin Elmer). Lysates were

35

analyzed by SDS-PAGE. Immunoblotting was completed with an anti-polyQ antibody (Chemicon) and visualized by chemiluminescence. Equivalent loading was confirmed by Coomassie staining.

5   Example 7. Alternative Loop Compatibility

Oligonucleotides encoding shRNAs directed against EGFP were obtained from Proligo (Boulder, CO). The loop structure was modified to either

5'-TCCAAGAGA-3'	SEQ ID NO:26
5'- CGTCTATATCATGGCCGACTCCAAGAGAGTCGGCCATGATATAGACGTTTTTT-3'	SEQ ID NO:27
5'-AATTAAAAAACGTCTATATCATGGCCGACTCTCTTGGAGTCGGCCATGATATAGA-CGGGCC-3'	SEQ ID NO:28
5'-TTCAAGAAA-3'	SEQ ID NO:29
5'-CGTCTATATCATGGCCGACTTCAAGAAAGTCGGCCATGAT ATAGACGTTTTTT-3'	SEQ ID NO:30
5'-AATTAAAAAACGTCT ATATCATGGCCGACTTTCTTGAAGTCGGCCATGATA-GACGGGCC-3'	SEQ ID NO:31

10   The oligonucleotides were annealed and placed into the pSilencer vector (Ambion) according to the manufacturer's protocol. All vectors were sequence verified. HeLa cells were transfected with p65Q and pAntiGFP, pTaiI, pTC-TaiI, pTT-TaiI, or empty pSilencer. Transfections were completed with Lipofectin™ and PLUS™ reagent (Invitrogen). After 48 hours, cells were lysed with RIPA buffer. Fluorescence was measured on a Victor V^2™ plate
15   reader (Perkin Elmer). Lysates were analyzed by SDS-PAGE. Immunoblotting was completed with an anti-polyQ antibody (Chemicon) and visualized by chemiluminescence. Equivalent loading was confirmed by Coomassie staining.

Example 8. Amplification of dsDNA ligated to UU-loop

20   Oligonucleotides were obtained from Proligo (Boulder, CO). The Flush Loop construct included part of the coding region for an shRNA against EGFP, derived from the sequence just 3' of the *Tai* I site in the EGFP cDNA.

5'-ATCCTCCAACGTCTATATCATGGCCGACTCCAAGAGAATACGTTGTTCGAG-CTACAACGTATTCTCTTGGAGTCCGGCCATGATATAGACGTTGGAGGAT-3'	SEQ ID NO:32

25   The structure that forms upon denaturation and cooling should be a stable blunt-ended hairpin. This was ligated with Cap2.

5'-CTGCCGAGTTCCTGCTTGAGATGCTGTTGAGUUACGTCGACTATCCTTGA ACACCAACTCGGCAG-3'	SEQ ID NO:33

36

The sequence of Cap2 is based on the cap sequence in the loop described in Kaur and Makrigiorgos (*Nucleic Acids Res.* 31:e26, 2003). Ligation reaction was completed with 100-fold excess of UU-loop using a rapid ligation kit (Roche). Reactants were subsequently digested with heat-labile uracil deglycosylase (Roche) for 10 minutes. PCR amplification was carried out with Titanium *Taq* per the manufacturer's protocol, using the following primers. The results were analyzed on a 2% agarose gel.

| Forward | 5'-CCTATAGTGAGTCGTATTACCTGCCGAGTT-3' | SEQ ID NO:34 |
| Reverse | 5'-CCTATAGTGAGTCGTATTACCTGCCGAGTT-3' | SEQ ID NO:35 |

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for producing a library of small hairpin RNAs (shRNAs), the method comprising:

providing a plurality of nucleic acid molecules comprising shRNA expression templates; and

transcribing shRNA from the shRNA expression templates, thereby producing a library of shRNAs.

2. The method of claim 1, wherein providing the plurality of nucleic acids comprising shRNA expression templates comprises:

providing a set of cDNA tags, each cDNA tag having a first end and a second end;

generating dual hairpin structures by ligating a first hairpin loop to the first end of each cDNA tag and a second hairpin loop to the second end of each cDNA tag, wherein the second hairpin loop comprises a cleavage site;

exposing the dual hairpin structures to an agent that cleaves the dual hairpin structures at the cleavage site to produce linearized, single-stranded dual hairpin structures; and

synthesizing, on the linearized, single-stranded dual hairpin structures, a second, complementary strand of DNA, thereby generating a plurality of shRNA expression templates.

3. The method of claim 1, wherein providing the plurality of nucleic acids comprising shRNA expression templates comprises:

providing a plurality of vectors each comprising a promoter and, operably linked to the promoter, an insert comprising the sense strand of a cDNA tag and a sequence that, when transcribed, will produce a hairpin loop;

transcribing the insert and extending the sequence by self-priming extension to generate a sequence complimentary to the sense strand of the cDNA tags, thereby producing a stem-loop-stem structure;

denaturing bonds between the nucleic acids along the stem of the stem-loop-stem structure to produce a denatured construct; and

synthesizing a second strand that is complimentary to the sequence of the denatured construct, thereby generating a plurality of shRNA expression templates.

4. The method of any of claims 1-3, wherein transcribing shRNA from the shRNA expression templates comprises:

modifying the shRNA expression templates by (a) operably linking the shRNA
expression templates to a promoter and (b), optionally, removing a portion of the sequence of
the first hairpin loop, thereby producing an shRNA expression construct and

transfecting cells with the shRNA expression construct.

5. The method of any of claims 1-3, wherein transcribing shRNA from the shRNA
expression templates comprises:

inserting the shRNA expression template into a plasmid vector, thereby producing an
shRNA vector construct;

optionally, removing a portion of the sequence of the first hairpin loop; and

transfecting cells with the shRNA vector construct.

6. The method of any of claims 1-5, wherein providing the set of cDNA tags
comprises:

providing a cDNA library;

exposing members of the cDNA library to at least two restriction enzymes,
wherein the restriction enzymes cleave the members of the cDNA library into fragments
10-50 nucleotides long.

7. The method of any of claims 1-5, wherein providing the set of cDNA tags
comprises:

providing a cDNA library, the members of which have been modified to include,
at a first terminus, an overhanging sequence representing a cleaved restriction site;

ligating members of the library to a first linker comprising (a) an overhanging
sequence complimentary to the overhanging sequence at the first terminus, thereby
reconstituting a first restriction site, and (b) a first immobilization agent, thereby
generating ligated members;

immobilizing the ligated members by exposing the first immobilization agent to a
substrate-bound partner, thereby generating ligated, substrate-bound members;

exposing the ligated, substrate-bound members to a first restriction enzyme that
cleaves the ligated, substrate-bound members at a second restriction site, thereby
generating a restriction-cut second terminus on the ligated, substrate-bound members;

exposing the ligated, substrate-bound members to a second restriction enzyme that
cleaves the first restriction site, thereby generating freed members;

ligating the freed members to a second linker comprising (a) an overhanging sequence complimentary to the restriction-cut second terminus, (b) a type IIS restriction site and (b) a second immobilization agent, thereby generating second linker-bound members;

exposing the second linker-bound members to a restriction enzyme that recognizes the type IIS restriction site and, by cleaving the second linker-bound members, generates a new first terminus;

immobilizing the second linker-bound members by exposing the second immobilization agent to a substrate-bound partner, thereby generating final substrate-bound members;

exposing the final substrate-bound members to the first restriction enzyme, thereby generating the set of cDNA tags.

8. The method of claim 6 or claim 7, wherein the cDNA library contains one or more members having sequences that are incompletely known.

9. The method of claim 6 or claim 7, wherein the cDNA library is normalized or subtracted.

10. The method of any of claims 6-9, wherein the cDNA library is obtained from a mammalian tissue.

11. The method of claim 10, wherein the mammalian tissue is a human tissue.

12. The method of claim 11, wherein the human tissue consists of a single organ or cell type.

13. The method of any of claims 1-12, wherein the shRNA library comprises at least 100 members.

14. The method of any of claims 2-13, wherein the cleavage site comprises a pair of uracil ribonucleic acids and the agent that cleaves the dual hairpin structures is uracil glycosylase or a biologically active variant or fragment thereof.

15. The method of any of claims 2-13, wherein the cleavage site comprises a restriction enzyme recognition site and the agent that cleaves the dual hairpin structures is a restriction enzyme.

16. The method of any of claims 2-15, further comprising the step of heating the dual hairpin structures after exposing the dual hairpin structures to an agent that cleaves the dual hairpin structures, wherein the heating is sufficient to denature the dual hairpin structures.

17. The method of any of claims 3-16, wherein the vectors are plasmids.

18. The method of any of claims 3-17, wherein the promoter is an inducible promoter.

19. The method of any of claims 3-17, wherein the promoter is an RNA polymerase III, RNA polymerase II, U6, T7, SV40, or H1 promoter.

20. The method of any of claims 3-19, wherein the vectors each comprise two promoters, one oriented on each side of the insert.

21. The method of any of claims 7-20, wherein the first restriction site is a *BamH* I restriction site.

22. The method of any of claims 7-21, wherein the first immobilization agent or the second immobilization agent is biotin or a polypeptide.

23. The method of claim 22, wherein the first immobilization agent is biotin and the substrate-bound partner is avidin or streptavidin or the first immobilization agent is a polypeptide and the substrate-bound partner is an antibody that specifically binds the polypeptide.

24. The method of any of claims 7-23, wherein the second restriction site is has a four base-pair recognition sequence.

41

25. The method of claim 24, wherein the restriction site is recognized by *Aci* I, *Alu* I, *Bfa* I, *BfuC* I, *BstU* I, *CviJ* I, *CviR* I, *Dpn* I, *Dpn* II, *Fat* I, *Hae* III, *Hha* I, *HinP1* I, *Hpa* II, *Mbo* I, *Mnl* I, *Mse* I, *Msp* I, *Nla* III, *Pho* I, *Rsa* I, *Sau3A* I, *Tai* I, *Taq$^\alpha$* I, *Tha* I, or *Tsp509* I.

26. The method of any of claims 7-25, wherein the type IIS restriction site is an *EcoP14* I, *Eco57* I, *Bpm* I, *BspH614* I, *Bco35* I, *Gsu* I, *Bce83* I, *Bsg* I, or *Mme* I restriction site, and the restriction enzyme that recognizes the type IIS restriction site is of *EcoP14* I, *Eco57* I, *Bpm* I, *BspH614* I, *Bco35* I, *Gsu* I, *Bce83* I, *Bsg* I, or *Mme* I, respectively.

27. The method of any of claims 1-26, wherein the step of transcribing shRNA from the shRNA expression templates is carried out within a cell.

28. The method of claim 27, wherein the cell is a cell in tissue culture.

29. The method of claim 28, wherein the cell in tissue culture is a primary cell.

30. The method of claim 27, wherein the cell is an animal cell or a plant cell.

31. The method of claim 27, wherein the cell is a cell *in vivo*.

32. The method of any of claims 27-31, wherein the cell is a human cell.

33. The method of any of claims 1-32, wherein the step of transcribing shRNA from the shRNA expression templates is carried out in a cell-free extract.

34. An shRNA library produced by the methods of any of claims 1-33.

35. A population of cells transfected with the shRNA library of claim 34.

36. A method of generating shRNA expression templates, the method comprising:

(a) providing a set of cDNA tags, each cDNA tag having a first end and a second end; generating dual hairpin structures by ligating a first hairpin loop to the first end of

42

each cDNA tag and a second hairpin loop to the second end of each cDNA tag, wherein the second hairpin loop comprises a cleavage site; exposing the dual hairpin structures to an agent that cleaves the dual hairpin structures at the cleavage site to produce linearized, single-stranded dual hairpin structures; and synthesizing, on the linearized, single-stranded dual hairpin structures, a second, complementary strand of DNA, thereby generating a plurality of shRNA expression templates; or

(b) providing a plurality of vectors each comprising a promoter and, operably linked to the promoter, an insert comprising the sense strand of a cDNA tag and a sequence that, when transcribed, will produce a hairpin loop; transcribing the insert and extending the sequence by self-priming extension to generate a sequence complimentary to the sense strand of the cDNA tags, thereby producing a stem-loop-stem structure; denaturing bonds between the nucleic acids along the stem of the stem-loop-stem structure to produce a denatured construct; and synthesizing a second strand that is complimentary to the sequence of the denatured construct, thereby generating a plurality of shRNA expression templates.

37. An shRNA expression template produced by the method of claim 36.

38. A vector comprising the shRNA expression template of claim 37.

39. A cell comprising the shRNA expression template of claim 37 or the vector of claim 38.

40. A composition comprising a plurality of cDNA tags.

41. The composition of claim 40, wherein the plurality comprises at least 100 cDNA tags.

42. The composition of claim 40 or claim 41, wherein the composition comprises at least one cDNA tag, the sequence of which is at least partially unknown.

43. The composition of claim 40 or claim 41, wherein the composition comprises at least one cDNA tag, the sequence of which is at least partially unknown.

43

44. A method of identifying a therapeutic target for the treatment of a disease, the method comprising:

    providing an shRNA library;

    providing a cell that serves as a model of the disease;

    contacting the cell with one or more shRNAs from the library; and

    evaluating an effect of the shRNA on a preselected parameter in the cell,

    whereby an improvement in the preselected parameter indicates that the shRNA identifies a therapeutic target.

44. The method of claim 43, wherein the cell is a simple organism.

45. The method of claim 43, wherein a single shRNA is expressed in the cell.

46. The method of claim 43, wherein a pool of shRNAs is expressed in the cell.

47. The method of claim 43, wherein the preselected parameter is selected from the group consisting of a metabolic parameter, a pathophysiological parameter, a developmental parameter, or a phenotypic parameter.

48. The method of claim 46, wherein the phenotypic parameter is selected from the group consisting of morphology, movement, and development.

49. The method of claim 46, wherein the pathophysiological parameter is selected from the group consisting of apoptosis, necrosis, proliferation, and senescence.

50. The method of claim 46, wherein the metabolic parameter is selected from the group consisting of enzyme activity; ion concentrations; ligand-receptor binding; and ligand-receptor activation.

51. The method of claim 50, wherein the change in enzyme activity is selected from the group consisting of a change in kinase, protease, helicase, or polymerase activity.

52. The method of claim 50, wherein the change in enzyme activity is selected from the group consisting of a change in ATP-dependent enzyme activity and ATP-independent enzyme activity.

53. The method of claim 50, wherein the change in ion concentrations is selected from the group consisting of a change in ion flux, ion gradient, plasma membrane potential, mitochondrial potential, and calcium concentrations.

54. The method of claim 43, further comprising identifying the shRNA.

45

2B.

2A.

2E.

2F.

3A.

Strand synthesis with biotinylated T7
primer and loop formation

pLoop vector

Loop

N_{16}

ACGT
TGCA

T7-Not I

N_{16}

ACGT

T7

Self-priming extension

3B.

Denature loops and
synthesize second strand

$T7_c$
T7

N_{16c}

ACGT
TGCA

N_{16}

N_{16}

ACGT
TGCA

T7
$T7_c$

N_{16c}

3C.

4.

mRNA

First strand synthesis and biotinylation

NotI-GA

RNase I treatment and retention of full-length RNA-DNA

NotI-GA

RNA hydrolysis and dG-tail

NotI-GA

Mix with biotinylated mRNA to remove abundant messages

NotI-GA

Normalized transcripts

Second strand synthesis with BamHI-dC primers

Abundant transcripts

BamHI-CCCCCCCC AAAAAAANotI
BamHI-GGGGGGGG TTTTTTTNotI-GA